DOCUMENT RESUME

ED 243 955 TM 840 271

AUTHOR Kohr, Richard L., Comp.; And Others
TITLE Getting Inside the EQA Inventory: Grades 5, 8, and 11. 1983 Edition.
INSTITUTION Pennsylvania State Dept. of Education, Harrisburg. Bureau of Educational Quality Assessment.
PUB DATE 83
NOTE 80p.; For related documents, see ED 178 591, ED 103 468, and ED 109 199.
PUB TYPE Guides - Non-Classroom Use (055) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Achievement Tests; Affective Measures; Affective Objectives; Basic Skills; Cognitive Objectives; *Educational Assessment; *Educational Quality; Elementary Secondary Education; Grade 5; Grade 8; Grade 11; Item Sampling; *State Programs; *Test Construction; *Testing Programs; Test Reliability; Test Validity
IDENTIFIERS *Pennsylvania Educational Quality Assessment

ABSTRACT

This guide begins with a series of questions and answers that introduce Pennsylvania's Educational Quality Assessment (EQA) Inventory as a 188- to 190-item multiple-choice test for grades 5, 8, and 11. Items are selected from a 400-item bank using matrix sampling procedures. Test results are analyzed at the school level; no individual student profiles are given. The 14 areas of measurement of the EQA are presented, with the measurement rationale and safeguards, and descriptions of (1) the assessment design, (2) the matrix sampling test design, (3) EQA reliability, and (4) EQA validity. The 14 areas are: (1) self-esteem, (2) understanding others, (3) reading comprehension, (4) writing skills, (5) mathematics, (6) interest in school and learning, (7) societal responsibility, (8) knowledge of law/government, (9) health, (10) creative activities, (11) career awareness, (12) appreciating human accomplishments, (13) knowledge of human accomplishments, and (14) information usage. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B. Ford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grades 5, 8 and 11

1983 Edition

Getting Inside the EQA Inventory

Compiled by Richard L. Kohr
Research Associate

James F. Hertzog
Basic Education Associate

and

Richard F. Seiverling
Basic Education Associate

PENNSYLVANIA DEPARTMENT OF EDUCATION
Bureau of School Improvement
Division of Educational Quality Assessment
Harrisburg, PA 17108
Revised Edition 1982
1983

Commonwealth of Pennsylvania
Dick Thornburgh, *Governor*

Department of Education
Robert C. Wilburn, *Secretary*

Office of Basic Education

Bureau of School Improvement
Donald L. Clark, *Director*

Division of Educational Quality Assessment
J. Robert Coldiron, *Chief*

Pennsylvania Department of Education
Box 911, 333 Market Street
Harrisburg, PA 17108

TABLE OF CONTENTS

Questions and Answers about EQA Inventory 1

Section One: Measurement Rationale 4

Self-Esteem 4

Understanding Others 6

Communication Skills–Reading Comprehension 8

Communication Skills - Writing Skills 9

Mathematics 10

Interest in School and Learning 11

Societal Responsibility 14

Knowledge of Law/Government 17

Health 19

Creative Activities 22

Career Awareness 25

Appreciating Human Accomplishments 27

Knowledge of Human Accomplishments 29

Information Usage 31

Section Two: Measurement Safeguards 33

DESIGNING AN ASSESSMENT PACKAGE 33

Overall Picture 33

A Word About Attitude 33

What About Paper-and-Pencil Measures of Attitudes? 33

Yes, But Can They Read It? 34

From Item Content to Answer Sheet 34

From Item Answers to Scores 35

Review of 1974-1977 Package 36

Review Committees 36

Content Revisions 37

Item Writing 39

Field Testing/Student Interviews 39

Item Analysis 40

Final Selection of Items 40

TABLE OF CONTENTS (Continued)

MATRIX SAMPLING TEST DESIGN 41
Matrix Sampling Procedure 41
Rationale for Matrix Sampling 41
Assignment of Items to Forms 42
Form Equivalence 42
Estimating School Means 43
RELIABILITY 49
Internal Consistency 49
Stability of School Means 49
Standard Error of the Mean 50
Percentage of Omits 50
VALIDITY 55
What is Validity? 55
Development Procedures 55
Validity Studies 55
Abstracts of Validity Studies Conducted 57
Self-Esteem 57
Understanding Others 57
Communication Skills and Mathematics 58
Interest in School and Learning 59
Societal Responsibility 59
Health 60
Information Usage 61
Reading Passage Dependence Study 62
REFERENCES 66
APPENDIX: Individuals participating on Test Review/Advisory Committees 67

QUESTIONS AND ANSWERS ABOUT EQA INVENTORY

What information is contained in this booklet?

Section One discusses the 14 areas measured and the devices associated with each. Included in this section are detailed measurement rationales, scale descriptions, and specifications for scoring. Section Two describes (1) the safeguards used to produce tests of high quality, and (2) how the EQA inventory fares on these checks.

Why are there eleven forms of the inventory?

Previous assessments took about four hours to complete because students answered all questions on the 400-item test battery. By employing a technique called *matrix sampling*, EQA has cut the administration time but also increased the number of items. Matrix sampling requires students to respond only to a portion of the total battery. Each portion (or form) is balanced with respect to item content so that a total school score may be generated. Statistically, it gives one a valid assessment, but takes much less student time.

Who takes the test battery?

All 5th, 8th and 11th grade students at the participating schools are asked to complete the EQA Inventory appropriate for their grade level. Students in grade 5 complete either Form A, B or C. Grade 8 and 11 students complete either Form A, B, C, or D.

How long is the inventory?

Each form of the questionnaire is designed to be completed by the student in about two hours. For grades 8 and 11, the 190 items (188 items for grade 11) per form can be completed in two one-hour sessions separated only by a short break. For grade 5, the 209 items can be completed in two 1 1/2-hour sessions on two consecutive days.

What kinds of tests are in the inventory?

Multiple-choice instruments tap cognitive skill, achievement and awareness in the areas of reading, writing, mathematical and problem-solving skills, health, citizenship, vocational knowledge and identification of famous persons. Some questions ask about students' attitudes and interests.

What do the tests cover?

Along with communication skills (reading and writing) and mathematical skills, the various instruments examine health knowledge and behavior, feelings toward self and others, value placed on learning and human accomplishments, participation in creative activities, career awareness, problem-solving skills, knowledge of law/government and human accomplishments, and societal responsibility.

Do the tests completely cover the Twelve Goals of Quality Education?

No. The goals are very broad statements covering many related concepts under one umbrella. An inventory of 5,000 items probably could not measure the goals in their entirety. Section One in this manual describes in detail which aspects of the goals are measured. Future development of instruments will enable more complete coverage.

What types of scoring procedures are used?

For the achievement tests, simply the number of correct answers is counted. The attitude scales are scored in two ways. First, each set of response options is given a weighting consistent with its corresponding item's direction (i.e., I like school; strongly agree = 3, agree = 2, disagree = 1, and strongly disagree = 0). Item scores are then summed to form a composite score. This is called norm-referenced scoring. The second procedure classifies options into categories of favorable and nonfavorable. In the above example the *strongly agree* and *agree* choices are given a score of one, while *disagree* and *strongly disagree* are given a score of zero. Students choosing favorable responses on a simple majority of scale items meet the criterion of minimum positive attitude. This technique is called criterion-referenced scoring.

Are the tests reliable?

Extensive investigation concerning the consistency of student responses within each scale (internal consistency) have been conducted by Division of Educational Quality Assessment personnel. All the scales demonstrate high internal consistency reliability. Reliability statistics for EQA instruments are presented in Section Two.

Are the tests valid?

Results for a group of studies conducted by the Division of Research lend validity support. Findings relating to test validity are presented in Section Two.

Are individual student profiles provided?

No. The unit of analysis of all data received from the Educational Quality Assessment Program is the *school.* No individual student profiles are given. In fact, students' names are not even placed on the answer sheets, thus insuring complete confidentiality of their answers.

Is the EQA Inventory the only source of information for the Educational Quality Assessment Program?

No. In addition to the student questionnaires, there is a survey of teachers and other information is collected from school and department records. These data are combined to generate a report for each participating school. For a complete description of the collection of this data, refer to *Manual for Interpreting School Reports.*

What kinds of information does the inventory provide?

Information includes (1) student-body standing on each area relative to a statewide-reference group, (2) student-body standing relative to groups similar in home and school environments, (3) proportion of student-body who demonstrate minimum positive attitudes, (4) proportion of student-body responses on the items in the battery, and (5) distribution of student-body scores across the range of form scores.

Are teachers held accountable for poor scores?

No. The Educational Quality Assessment Program uses three separate assessment inventories to examine student goal achievement at grades 5, 8 and 11. Students at other grade levels do not take the tests. Test outcomes are not solely a result of what teachers at those three levels are or are not doing. Student attitudes and achievements are a complex product of the total home, school and community experience. Accountability only comes into play in terms of taking quality assessment results into consideration when trying to meet the needs of students.

Does the PDE offer help to improve school programs?

Yes. PDE curriculum specialists constantly investigate new approaches and review research literature for schools which request help to improve their programs.

Is there any indication that schools can improve student attitudes by implementing programs?

Yes. Several programs developed by school districts have already yielded measurable improvements on the EQA Inventory scales.

The Division of Educational Quality Assessment has published in limited quantities a series of resource packets related to each of the areas measured in the assessment package. The purpose of these packets is to guide the reader to a variety of sources of information, materials, consultants and adoptable programs which may effect improvement in a goal area identified critical by a school district.

Packets have been published in the following goal areas: Self-Esteem, Understanding Others, Reading, Writing Skills, Mathematics, Interest in School and Learning, Societal Responsibility and Citizenship, Health and Safety, Creative Activities, Career Awareness, Appreciating Human Accomplishments, and Information Usage.

For those participating districts wishing to avail themselves of these materials, it is strongly recommended that requests be limited exclusively to those packet(s) which relate to identified district priorities. Requests in writing should be initiated on district letterhead to the Division of Educational Quality Assessment, Pennsylvania Department of Education, Box 911, 333 Market Street, Harrisburg, PA 17108.

Because of severe printing limitations in the PDE, only one request for any given packet will be honored; however, the materials in the packet(s) may be duplicated locally without special permission.

Note: These resource packets supersede the previously published booklet, "What Some Pennsylvania School Districts Are Doing To Raise Their EQA Scores," which publication is out of print. Sections of the newly published resource packets incorporate the identical materials previously contained in the publication herein mentioned.

Section One: Measurement Rationale

Self-Esteem

MEASUREMENT RATIONALE

Students, regardless of their level and pattern of particular talents, should experience a school setting that will strengthen, not damage, their self-esteem. The school should help students at all levels of talent achieve feelings of success and of acceptance that are important in relating appropriately to others and in becoming a successful member of society.

Self-esteem is a personal judgment of worthiness. It is a subjective experience which the individual conveys to others verbally or by other behavior. Most theories acknowledge that our self-image and feelings of worthiness are determined largely by how well we can live up to our own aspirations and meet expectations of others.

Aspirations become closely associated with personal goal-setting behavior originating in our internalized system of values. Expectations are external in nature and are related to goals set collectively by society or by significant individuals in our lives. Assessment in this area is based on several components believed to be related to the development of positive self-esteem in the school context.

The first is a feeling of acceptance. The feeling of self-worth and acceptance by teachers and classmates is an important ingredient to the development of a positive self-esteem. Those lacking a feeling of acceptance often have difficulty in interpersonal relations and tend to believe that others have little confidence in, or low regard for, them.

Secondly, feelings of self-determination and success in school work, class recitation and teacher relationships help to mold feelings of self-confidence – so important to adequate coping in daily life. Those who lack feelings of success and confidence are often characterized as cautious, submissive individuals who feel inadequate, are fearful and expect to be unsuccessful in dealing with new situations.

GENERAL SCALE DESCRIPTION*

For each grade level the self-esteem scale is comprised of 32-33 short, self-description statements. For grade 5, 15 items are positively worded – describing the student in a favorable light – and 18 are negatively worded, characterizing the student in a negative vein. The grade 8 items include 14 positively worded items and 18 negatively worded items. For grade 11, 13 items are positively worded and 19 are negatively worded.

Sample positively worded item: **I get along well with kids in my class.**

Sample negatively worded item: **I feel unhappy about my schoolwork.**

Response options available to the students are: (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree, and (4) Strongly Disagree.

For norm-referenced scoring, the item weighting scheme used is:

Response Choices

Item Direction	Strongly Agree	Mostly Agree	Mostly Disagree	Strongly Disagree
Positive	3	2	1	0
Negative	0	1	2	3

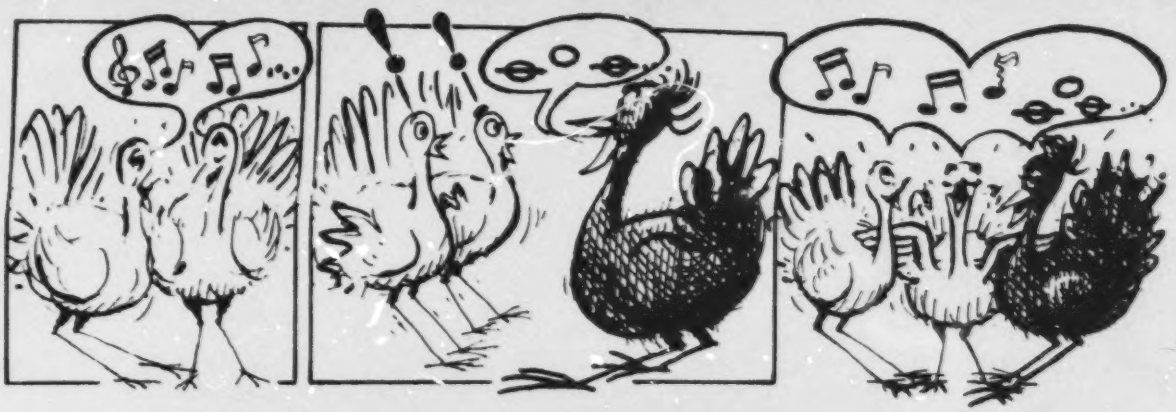
For criterion-referenced scoring, responses are considered *favorable* if they reflect a positive self-image. An individual's score on the scale is the percentage of items to which a favorable response was given. For the self-esteem instrument the scoring scheme applied to the items is:

Response Choices

Item Direction	Strongly Agree	Mostly Agree	Mostly Disagree	Strongly Disagree
Positive	1	1	0	0
Negative	0	0	1	1

---

*The self-esteem scale is a result of extensive revision of the Self Esteem instrument developed by Stanley Coopersmith. Richard L. Kohr and Dennis Deck from the Division of Educational Quality Assessment were responsible for the revisions.



Understanding Others

MEASUREMENT RATIONALE

Students should acquire attitudes which will facilitate easy interaction with all people regardless of skin color, religious beliefs, physical characteristics, socioeconomic status or cultural tradition. Schools, although not solely responsible for the development of these attitudes, should be aware of the degree of tolerance exercised by their students and should promote positive interpersonal relations in the school setting.

Each individual daily experiences a variety of types of social interactions. One way of categorizing these types of interactions is into four types of social interactions that exist: (1) Friendship, (2) Social Distance, (3) Respect, (4) Superordination.

An assumption made in assessing this goal is that tolerance toward differing others is to a degree situational. The instrument employed contains items to measure tolerance in the four types of situations described above.

Thus, it was assumed, for example, that whereas an individual might have no difficulty accepting someone different from himself or herself in situations which could be categorized as social distance (e.g., sitting beside the person in class), the same individual might not want to serve on a committee headed by this person (superordination).

Also, an individual might not be willing to spend a great deal of time interacting socially with someone different from himself or herself (friendship), but might have little reluctance in asking the person's opinions about an important issue (respect).

GENERAL SCALE DESCRIPTION*

Students are presented with descriptions of school situations involving interactions with others who differ from themselves. Below each situation are descriptions of five types of students, each of whom differs from the student responding in one of the following five ways: economic status, religion, race, physical appearance or cultural tradition. Depending upon the type of situation presented, the student is asked either whether he or she would want to interact with each of

---

*The Understanding Others instrument was developed by James Masters, James Dorwart and George Brehman from the Division of Research.

the differing others, or whether he or she would mind interacting with each of them. For situations in which the student is asked whether he or she would want to interact with each of the other students, some motivation for desiring this has been given (e.g., Each of the students is an expert at a skill he or she would like to learn). For situations in which the student is asked whether he or she would mind interacting with each of the students, he or she has been given no choice in placing himself or herself in the situation (e.g., Each student must share a locker with another student).

The grade 5 instrument includes a total of 45 items; those for grades 8 and 11 each contain 40 items. The response options available to students at all three grade levels are: (1) Yes, (2) I think so, (3) I don't know, (4) I don't think so, and (5) No.

Sample Item: **Would you mind sitting in class beside a student who:**

Has few friends because of the way he or she looks?

Is very much like you, but whose skin color is different from yours?

Is much poorer than you?

Belongs to a religion which you do not belong to?

Is from another country and doesn't speak English very well?

The structure of the instruments makes it possible to examine item data more closely by grouping it in one of two ways, either by type of situation or by type of differing other. The grade 5 instrument contains 15 friendship items and 10 items for each of the interaction categories of social distance, superordination and respect. The instruments for grades 8 and 11 contain 10 items of each of the four types of interaction. Grouping the items of the grade 5 instrument by type of differing other results in 9 items for each of the five types (e.g., race, religion). The same process for the instruments for grades 8 and 11 produces 8 items for each of the five types.

For norm-referenced scoring, the following weighting scheme is used:

Response Options

Question Type	Yes	I Think So	I Don't Know	I Don't Think So	No
Would you want to . . .?	4	3	2	1	0
Would you mind . . . . . ?	0	1	2	3	4

For criterion-referenced scoring, responses are considered *favorable* if they reflect acceptance when interacting with differing others. An individual's score on the scale is the percentage of items to which a favorable response was given. The scoring scheme is:

Response Options

Question Type	Yes	I Think So	I Don't Know	I Don't Think So	No
Would you want to . . .?	1	1	0	0	0
Would you mind. . . . .?	0	0	0	1	1

Communication Skills—Reading Comprehension

MEASUREMENT RATIONALE

The goal of all reading is the comprehension of meaning. Comprehension involves the integration of a complex set of behaviors which begins with the association of a sound with a given symbol and moves through word, sentence and paragraph meaning to higher level passage-interpretation skills. Students should master such fundamental skills as word attack, phonetic analysis and structural analysis at their level of sophistication, but a general reading assessment should focus on the total reading process by means of some measure of reading comprehension.

GENERAL SCALE DESCRIPTION

Measuring literal and inferential comprehension tasks indirectly assesses the fundamental reading skills and is perhaps the best global measure of reading achievement in a survey form with limited space. Comprehension tasks at various difficulty levels can be delineated, and the 48 items in the scales at each grade level measure some of them:

Level	Comprehension Tasks
Literal: Comprehension of details, relationships and major ideas which are explicitly stated in a passage	Identifying main idea Identifying supporting details Identifying cause and effect
Inferential: Comprehension of details, relationships and major ideas which are implied but not explicitly stated in a passage	Inferring main idea Using supporting details Inferring cause and effect Predicting outcomes Drawing conclusions Determining author's point of view Making judgments Determining the [appropriate] meaning of a [multiple-meaning] word/contextual analysis

Reading consultants and staff testing specialists constructed the test items. The passages on which the items are based are either narrative or expository. For example, excerpts from poetry, literature, news and weather reports, and a warranty serve as passage material. Each item was reviewed to insure that it was passage dependent (could not be answered unless the student had read the passage) and that it was not biased against a cultural group or against either sex.

The scales are scored on the basis of one point for each correct answer. No adjustment is made for guessing.

Communication Skills—Writing Skills

MEASUREMENT RATIONALE

Communication skills include the ability to express ideas through writing. To measure writing skills there are two appropriate procedures: an essay test which requires the student to write several short essays, letters, and reports – or an objective test which requires the student to choose the best way of expressing something. While the essay test engages the student in actual writing, it is very time consuming and costly to administer and score.

The objective test for writing skills used in this program measures some of the skills involved in writing standard edited English. The skills include punctuation, capitalization and usage. Also included are standard forms of organization and development of sentences and paragraphs, and elements of style, tone and flavor.

With carefully designed items assessing this broad range of skills, this type of measurement is highly correlated with tests of actual writing. Research has shown that students who do well on such a test tend also be competent writers.

Four categories of skills are assessed:

1. *Mechanics and usage* - recognizing the application of conventions
2. *Sentence sense* - recognizing clear, concise and effective sentences
3. *Paragraph sense* - recognizing coherence, unity and transition in the organization and development of paragraphs and larger units
4. *Style, tone and flavor* - recognizing effective and appropriate language for a given purpose, audience or situation

GENERAL SCALE DESCRIPTION*

For grade 5, 45 items measure the four skill categories above. At grades 8 and 11, there are 60 items per scale.

The scales are scored on the basis of one point for each correct answer. No adjustment is made for guessing. Sample item: **Choose the answer that is a complete sentence.**

A. Happily skipping down our street.
B. Singing happily, I skipped down our street.
C. Singing and skipping happily down our street.
D. Skipping down our street and singing happily.

---

*These scales were developed by Dennis Deck and John Meehan of the Department of Education with the aid of an advisory committee of writing specialists.

Mathematics

MEASUREMENT RATIONALE

The mathematics test is designed to measure the skills and concepts that are generally considered to be important for all students. The set of skills and concepts were identified using the PRIMES* Content Authority List. Items measure the concept itself, computation related to the concept or problem solving involving the concept.

GENERAL SCALE DESCRIPTION**

The mathematics scales are 60-item tests. Items include questions about: number systems, numeration and notation, geometry, measurements, number patterns and relationships, and other topics.

The scales are scored on the basis of one point for each correct answer. No adjustment is made for guessing.

Sample items:

Grade 5

How many ounces are there in a pound and a half?

A 10 2/3 ounces
B 22 ounces
C 24 ounces
D 32 ounces

Grade 8

57.9 - 13.51 =

A 43.39
B 44.41
C 44.39
D 44.49

Grade 11

If 29 students in a senior class of 232 own cars, then the ratio of car owners to non-car owners is:

A 7:1
B 1:8
C 1:7
D 8:1

There are no higher level concepts such as those encountered in trigonometry, calculus or even high school algebra. Measurement and estimation using both the English system and the metric system are included.

---

* The Pennsylvania Retrieval of Information for Mathematics Educational System–a PDE project.
** These scales were developed by Carl Guerriero, Joyce Kim and Richard Kohr of the Division of Educational Quality Assessment with the aid of an advisory committee of mathematics specialists.

Interest in School and Learning

MEASUREMENT RATIONALE

The school represents perhaps the most powerful single force in determining a person's overall attitude toward learning. The climate and learning atmosphere in the school, the educational experiences the school provides and the quality of the personal interactions it fosters between student and educator all shape the students' life-long attitudes toward learning.

The school experience should be such that students find the learning activities associated with it enjoyable and rewarding to the point that they are motivated to do well and to continue learning on their own initiative beyond the requirements of formal education. Everything possible should be done to ensure that the attitude of the teacher, the atmosphere of the school and the school's physical condition contribute toward this end so that the individual – both as a child and later as an adult – will hold education high among his or her values.

The schools that pupils attend influence both their present attitudes toward learning and their attitudes toward learning throughout their lives. Both subject matter being learned in school and the environment in which this learning takes place are important factors in forming lasting impressions about learning. Thus, the instrument used to measure this goal includes items concerned with attitudes toward the school environment, as well as items concerned with attitudes toward learning.

GENERAL SCALE DESCRIPTION* – Grade 5

The interest in school and learning scale includes 27 items, all of which are stated as questions. Pupils are asked how they feel about their schoolwork, about their school and their teachers, and about learning in general. The response options available to the pupil are: (1) Very Happy, (2) A Little Happy, (3) A Little Unhappy, and (4) Very Unhappy.

The items within the scale cover two areas: **Attitude toward learning** items measure pupil attitudes toward learning in or out of school and toward school subjects. Sample items: *How do you feel when you learn about science in school? How do you feel when you are given a book for a birthday present?*

School climate items measure pupil attitudes toward their school environment and toward their teachers and principals. Sample item: *How do you feel when you talk to your teacher?* All items except one are positively worded.

For norm-referenced scoring, the following weighting scheme is used:

	Response Options			
Item Direction	**Very Happy**	**A Little Happy**	**A Little Unhappy**	**Very Unhappy**
Positive	3	2	1	0
Negative	0	1	2	3

For criterion-referenced scoring, responses are considered *favorable* if they reflect student pleasure with positive statements about school and learning or displeasure with negative statements concerning school and learning. A student's score on the scale is the percentage of items to which a favorable response was given. For this scale, the scoring scheme applied to the items is:

	Response Options			
Item Direction	**Very Happy**	**A Little Happy**	**A Little Unhappy**	**Very Unhappy**
Positive	1	1	0	0
Negative	0	0	1	1

GENERAL SCALE DESCRIPTION* – Grade 8

This scale is made up of 40 statements about the school, teachers, course content and the learning experience. Twenty-six of the items are favorable statements about school and learning, e.g., "Most of my subjects this year are worthwhile." The remaining items are negatively stated, e.g., "The teachers in this school do *not* give enough praise when you do well." Response options available to the student are: (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree, and (4) Strongly Disagree.

The items within the scale can be grouped into two 20-item sets:

School learning measures student feelings about their courses, their assignments and their teachers' methods. Sample item: *Most of my subjects this year are worthwhile.*

School environment measures student feelings about the climates of their schools, including perceptions of teacher-student relations. Sample item: *The school principal really seems to care about students.*

For norm-referenced scoring, the following weighting scheme is used:

	Response Options			
Item Direction	**Strongly Agree**	**Mostly Agree**	**Mostly Disagree**	**Strongly Disagree**
Positive	3	2	1	0
Negative	0	1	2	3

For criterion-referenced scoring, responses are considered *favorable* if they reflect student agreement with positive statements about school and learning or disagreement with negative statements concerning school and learning. A student's score on the scale is the percentage of items to which a favorable response was given. For this scale, the scoring scheme applied to the items is:

Response Options

Item Direction	Strongly Agree	Mostly Agree	Mostly Disagree	Strongly Disagree
Positive	1	1	0	0
Negative	0	0	1	1

GENERAL SCALE DESCRIPTION*– Grade 11

This scale includes 40 statements about school, teachers and the learning experience. Twenty-nine items cast these in a favorable light, e.g., "Most of my teachers do a good job of teaching." The remaining items are stated negatively, e.g., "Students are under too much pressure in this school." Response options available to the student are: (1) Strongly Agree, (2) Mostly Agree, (3) Mostly Disagree, and (4) Strongly Disagree.

The total scale is comprised of three areas:

Learning items to measure students' feelings about learning, both in their present schools and in general. Sample item: *I like learning social studies in this school.*

School environment items to measure students' general feelings about school, including their feelings about the way students are treated in their schools and the value of school to them. Sample item: *There are interesting activities to look forward to in this school.*

Teacher-Student relations items to measure students' feelings about their teachers. Sample item: *My teachers are interested in how well I do my assignments.*

For both norm-referenced and criterion-referenced scoring, the weights are the same as those for grade 8 listed above.

---

*The Interest in School and Learning scales were developed by James Masters, James Dorwart, Richard Kohr and Dennis Deck from the Department of Education.

Societal Responsibility

MEASUREMENT RATIONALE

Teachers, administrators, curriculum developers, parents and other concerned citizens should encourage students to assume responsibility for their actions and the actions of groups to which they belong. Opportunities should be provided for students to work cooperatively and demonstrate integrity when dealing with others. Students should be given the chance to take the initiative and assume leadership as well as support group efforts as followers.

A person's behavior may change under differing sets of motivating conditions. A behavior seen as unacceptable according to tradition or law may be seen as acceptable to an individual under a certain set of conditions.

When behaviors which infringe on the rights and property of others, which ignore the representative government which has established the laws, and which abandon one's personal development are undertaken to points exceeding the norm and points beyond the criterion set by society, the school should be aware of the situation before a pattern of continually relaxing motivators sets in.

To assess societal responsibility, varying situations are presented with conditions of reward and punishment. Students indicate whether they are willing to continue with positive behaviors under each of the conditions.

GENERAL SCALE DESCRIPTION* - Grade 5

This scale contains 27 items which ask students whether they think they will take a given action in a specific situation. Items are couched as *if* statements, e.g., *If a store clerk gave me too much change, I would return the extra money.* The prestated actions in some of the items reflect irresponsibility. Other items measure students' willingness to display responsible behaviors. Response options are: (1) Yes, (2) Maybe, and (3) No.

The items can be grouped into three item groups:

Concern for the welfare and dignity of others items measure concern for the feelings of others, willingness to accept new people into a group, and tendency to go to the aid of others in distress.

Also measured is the degree of restraint from teasing or degrading others. Sample item: *If someone in my class wore odd clothes, I would tease the person about it.*

Respect for law and authority items measure the willingness to report law-breaking of others, obey authorities during emergencies and prevent classroom disruptions. Also assessed is the degree of restraint from actions that could harm others or damage property. Sample item: *If everyone else was breaking a school rule, I would also break it.*

Personal responsibility and integrity contains items which tap the willingness to honor self-made commitments to individuals or groups and the willingness to take responsibility for one's own mistakes and to report mistakes made in one's favor. Sample item: *If I was "fooling around" with the school's record player and broke it, I would tell the teacher.*

The following item weighting scheme is used for norm-referenced scoring:

Behavior Direction	Yes	Maybe	No
Positive	2	1	0
Negative	0	1	2

For criterion-referenced scoring, responses are considered *favorable* when they reflect a willingness to display responsible behaviors or an unwillingness to use irresponsible behaviors. A student's score on the scale is the percentage of items to which a favorable response was given. For this scale, the scoring scheme applied to the items is:

Behavior Direction	Yes	Maybe	No
Positive	1	0	0
Negative	0	0	1

GENERAL SCALE DESCRIPTION* - Grades 8 and 11

Forty-eight items for grade 8 and 36 items for grade 11 measure willingness to exhibit responsibility in many social situations under a variety of motivating conditions. Social contexts are given by situations posing problems and suggesting actions predefined as acceptable or unacceptable behavior. Given several positive or negative consequences resulting from the action, students are asked to decide whether to take the action for each consequence.

Sample story (grade 8):

A student accidentally breaks a school window. If I was that student, I would TELL THE PRINCIPAL OR TEACHER about my breaking the window. . .

Sample item set:	Yes	Maybe	No
even if I would have to pay for the window.	Y	M	N
since it is the honest thing to do.	Y	M	N
if the entire class would be punished if the person who broke the window were not caught.	Y	M	N

Sample situation (grade 11):

A student gets a dog. A city law requires that a $10 dog tag be bought by the dog's owner. In this situation, I would BUY THE TAG. . .

Sample item set:	Yes	Maybe	No
if the fine for not having a tag was $25.	Y	M	N
if the dog might be more easily lost or stolen without a tag.	Y	M	N
since owners are responsible for their dogs.	Y	M	N

The items within each scale can be grouped into the three areas listed and described above under grade 5 description.

The following item weighting scheme is used for norm-referenced scoring:

	Response Options		
Direction	**Yes**	**Maybe**	**No**
Acceptable Behavior	2	1	0
Unacceptable Behavior	0	1	2

For criterion-referenced scoring, responses are considered *favorable* when they reflect a willingness to display proper behaviors or an unwillingness to use poor behaviors. A student's score on the scale is the percentage of items to which a favorable response was given. The scoring scheme applied to the items is:

	Response Options		
Direction	**Yes**	**Maybe**	**No**
Acceptable Behavior	1	0	0
Unacceptable Behavior	0	0	1

---

*Nolan Russell and Richard Kohr from the Division of Educational Quality Assessment originally developed these scales. George Brehman, Gregory Shannon and Carl Guerriero from the Department of Education revised the scales for the current battery.

Knowledge of Law/Government

MEASUREMENT RATIONALE

It must be recognized that a major prerequisite to developing habits and attitudes must be the acquiring of basic knowledge. Obviously, students who have gained knowledge about our system of government will not automatically become responsible citizens. However, the development of responsible citizenship behaviors will be impaired if certain basic knowledges are not obtained. For example, it would be expected that students who do not understand what voting is all about would not exercise this right when out of school. Students who do understand the voting process may or may not vote, but at least their decision will not be based upon ignorance.

GENERAL SCALE DESCRIPTION*

In consideration of this, three major areas of knowledge were identified as essential to good citizenship. They are broadly described as knowledge of government, knowledge of law, and knowledge of interdependence of people.

Subsumed under government is basic knowledge about the structure and functions of government. As stated in the National Assessment of Educational Progress publication describing the citizenship objectives they measure regularly on a national basis: "Citizens must be familiar with the working of government and the channels through which they can exert influence if they are to participate effectively in governmental processes."

In order for students to become responsible citizens they must develop a respect for the law, a feeling that it serves them as well as restricts them. These attitudes cannot be formed without an understanding of why laws are needed, how they are made and how citizens can affect the lawmaking process. Also, students must develop an understanding of how the law relates to them as individuals.

Knowledge of interdependence of people items include the measurement of knowledge about how people can most effectively interrelate. To function most effectively as citizens, students need to understand how their day-to-day actions affect others, how conflicts occur and how they can be resolved. In addition, as the population of the world increases and vital resources become scarce, these concepts become increasingly important for citizens to apply when thinking about our country's relationships with the rest of the world.

Sample items:

The main job of Congress is to

A make sure laws are obeyed.
B make laws.
C decide whether someone has broken the law.
D make everybody happy with the laws.

For a person to get a law changed, it is illegal to

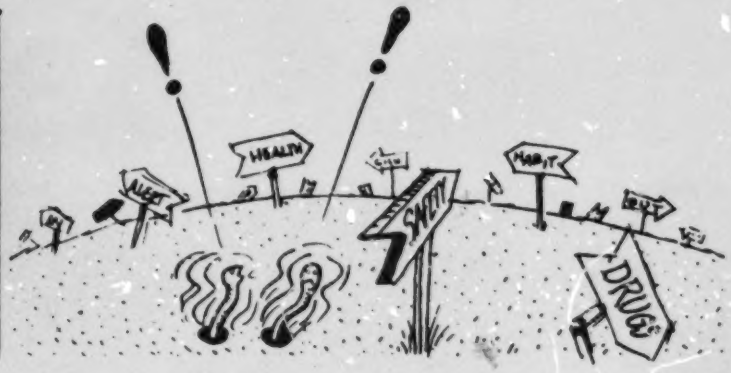
A join political action groups.
B lobby.
C pay someone to have it changed.
D make campaign contributions.

A major purpose of the United Nations is to

A help nations settle arguments without fighting.
B organize student exchange programs between nations.
C replace nations with a world government.
D spread democratic government to all nations.

The scales are scored on the basis of one point for each correct answer. No adjustment is made for guessing.

[9] These scales were developed by Eugene Skiffington, Robert Schell, James Masters, James Dorwart and Francis Reardon from the Department of Education with the aid of an advisory committee of social studies experts.



Health

APPROACH

In their own interest, as well as the interest of society at large, students should know how to take care of themselves and how to keep physically fit. They should know what the requirements are for physical and mental health and what practices harmful to health should be avoided.

After gaining this knowledge, they should acquire habits which increase the probability of remaining healthy and fit throughout life. In cases where the home has been deficient in encouraging the child to practice some health habits, the school has an obligation to be aware of the situation and to see that opportunities to remedy the deficiency are provided.

Knowledge of Health/Safety—Grade 5

MEASUREMENT RATIONALE

To develop sound health-promoting behavior, an individual must acquire a basic understanding of how the body develops and functions. Understanding disease processes, their development and prevention, as well as principles related to dental care, nutrition, personal hygiene, physical fitness, safety, drug use and general care of the body, are all instrumental in forming behavior patterns supportive of good health.

Assessment of this goal at the elementary grade level concentrates on the knowledge base believed to be representative of the general health domain.

GENERAL SCALE DESCRIPTION*

The grade 5 instrument contains 45 items which cover three main areas: sociological health, safety and personal health. The sociological health area concentrates on problems and effects of smoking, alcohol consumption and drugs. The safety area deals with proper practices relating to fire, electricity, bicycle riding, swimming, traffic, first aid and other situational contexts. Items in the personal health area pertain to dental care, nutrition, disease control, exercise, digestion and general care of the body.

---

*Susan Frey, James Dorwart, Dennis Deck, Richard Kohr, Douglas Boelhouwer and Francis Reardon from the Department of Education developed and refined this scale.

This instrument is regarded as an adequate group measure of knowledge in the general health area. A multiple-choice format is employed with four response options per item. The instrument is scored by assigning one point to each correct answer. No adjustment is made for guessing.

Sample items:

Which of the following is a safety rule to follow when riding a bicycle?

A Ride on the right side of the road.
B Ride on the left side of the road.
C Ride on the sidewalk.
D Wear dark-colored clothing when you ride.

A balanced diet includes foods that are

A digested quickly and easily.
B high in proteins and vitamins.
C from each of the four basic food groups.
D low in calories and fats.

If you burn yourself, you should first

A put butter on the burn.
B run cold water on the burn.
C cover the burn tightly with a clean bandage.
D put nothing on the burn.

Health and Safety Practices—Grades 8 and 11

MEASUREMENT RATIONALE

Understanding how diseases and their prevention, dental care, nutrition, personal hygiene, physical exercise, safety and drug use relate to the structure and function of the human body is an important first step in each individual's willingness to consistently exhibit habits which are conducive to the maintenance of personal health and well-being.

Therefore, assessment in this goal area attempts to get at students' willingness to display proper health behaviors in a variety of situations.

GENERAL SCALE DESCRIPTION*

The student is asked to decide whether he or she would take a given health-related action. Each action is predefined as indicating either good or poor health practice. Stimulus contexts surrounding the choices are systematically varied.

The format of each question is a situational story about a make-believe high school student. The respondent is first asked to consider taking a specific action. In each question three motivation-including conditions, i.e., rewards and punishments, are made contingent upon the taking of the action.

---

*The health habits scales were originally developed by Nolan Russell, Division of Educational Quality Assessment. Revisions for the current battery were made by Susan Frey, Dennis Deck and Richard Kohr.

Sample story:

After school, I am talking with friends and someone offers me a cigarette. I would TAKE THE CIGARETTE. . .

Sample item set:

	Response Options			
	Definitely Would	**Probably Would**	**Probably Would Not**	**Definitely Would Not**
if the other kids were smoking.	A	B	C	D
if I thought that my parents would not find out.	A	B	C	D
even though I know it could lead to heart or lung disease.	A	B	C	D

The items within each scale can be grouped into three areas:

Personal and community health includes willingness to follow proper diet, to take proper medical precautions, to use good personal hygiene practices and to refrain from interpersonal contacts when ill.

Personal and community safety measures the degree of restraint from unnecessary risk-taking at home, at school and at play and restraint from submitting others to undue risks.

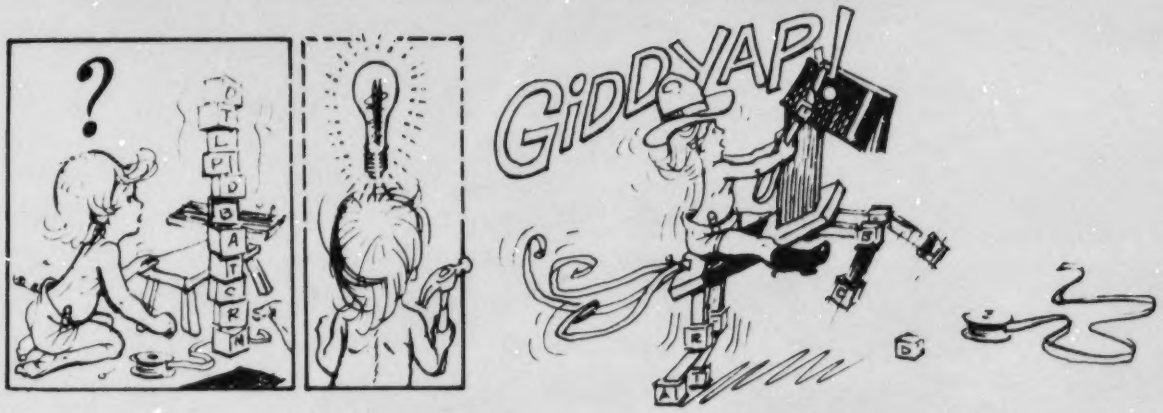
Drugs measures restraint from (1) improper use of prescription drugs, (2) experimentation with drugs, and (3) maintaining close contact with others who are using drugs. Improper use of prescription drugs includes restraint from using old medicine, medication prescribed for others, or more medicine than has been prescribed by the doctor.

For norm-referenced scores, the following item weighting scheme is used:

	Response Options			
Behavior Direction	**Definitely Would**	**Probably Would**	**Probably Would Not**	**Definitely Would Not**
Good Health Practice	3	2	1	0
Poor Health Practice	0	1	2	3

For criterion-referenced scoring, responses are considered *favorable* when they reflect a willingness to take good health-related actions or an unwillingness to display behaviors that might be harmful to health. A student's score on the scale is the percentage of items to which a favorable response was given. For the health scale, the scoring scheme applied to the items is:

	Response Options			
Behavior Direction	**Definitely Would**	**Probably Would**	**Probably Would Not**	**Definitely Would Not**
Good Health Practice	1	1	0	0
Poor Health Practice	0	0	1	1



Creative Activities

MEASUREMENT RATIONALE

The notion of creativity has been variously defined. Here it encompasses worthwhile activities that a student initiates and pursues on his/her own – activities having an outcome that is perceived by the student and by others as a contribution to some part of his/her world. Such activities can be included in a wide variety of fields, not only the sciences and the arts, but also the organization of human affairs and the development and exercise of salable skills in the production of any of a host of practical things that enrich our way of living.

It is not to be expected that the creative behavior of students will, except in very rare cases, lead to wholly original productions. What is expected is that many students, if given the chance, will become involved in exploring ideas and ways of doing things that are new and existing within their own world. The school should make possible this kind of rewarding experience for all students, and should judge its success by its ability to evoke in students a flexible and creative approach to human affairs.

To measure the two dimensions of the goal, i.e., the frequency of involvement in creative activities and the interest in future creative activity, a dual response set was placed on a creative activities checklist.

The first dimension is based on student experiences during the past three years to determine if the opportunities have been available and if the student has taken advantage of them. These opportunities are not limited to those provided within the school because: (1) it is not possible to isolate community and school functions, (2) the school often promotes and supports community activities, and (3) the amount of community opportunity is often the direct result of the educational background of that community.

The second dimension is based on the student expressed interest in future participation in creative activities.

GENERAL SCALE DESCRIPTION*

The creative activities checklists each contain activities (36 for grade 5 and 40 for grades 8 and 11) which have been judged to be creative on the basis of one or more of the following criteria:

a. *Newness of approach*- -The activity is one in which the student undertakes a purposeful search for a different pathway to the anticipated product than that which has been commonly followed in the past.

b. *Originality*- -The activity is one which results in an independent and/or spontaneous, firsthand production with a freshness of aspect or design or newness of style or character.

c. *Adaptiveness*--The activity is one which involves a modification of a method, artifact, or process to serve a different purpose, to fit a new set of requirements, or to produce a different product.

d. *Occupational Uniqueness*- -The activity is one which is not usually expected from the age level of the student.

e. *Evocation of Feeling and Communication of an Idea*- -The activity results in the production which transmits a feeling or an idea to another person or to other persons.

Each item describes the activity and asks the students to indicate the frequency of participation in that activity and the future interest in that activity.

To the question--"During the past three years, how many times have you done the following?"- -the student chooses for each activity: (1) Have not done, (2) Have done once, (3) Have done 2 or 3 times, (4) Have done about 5 times, or (5) Have done about 8 times (or more). Then, for each activity, the student chooses: (1) No time, (2) Very little time, (3) Some time, or (4) A lot of time- -for the question- -"In the future, how much time would you like to spend doing the activity?"

The scale for each grade level contains five item groups, each balanced equally.

1. **Visual arts** include sculpturing, painting, sketching, drawing, designing decorative and useful items, designing clothing, doing layout work for a publication and printmaking or graphic design.

2. **Performing arts** include activities dealing with singing, speech, dancing, music and acting.

3. **Writing** activities include writing poetry, news items, editorials, themes, essays, stories, plays, scripts, lyrics, speeches and reviews.

4. **Science** activities include experiments with objects or living things; designing small buildings, machines or displays; studying animals; working with electronic equipment; writing scientific papers; and applying scientific principles.

5. **Social and practical studies** include projects and displays, organization of groups, promotion of candidates or activities, written discussion of social issues, developing athletic routines, editing publications and developing recipes.

---

*The Creative Activities scale was developed by James Masters and James Hertzog from the Department of Education with the aid of many PDE curriculum specialists.

For norm-referenced scores, all items in this scale are positively worded and the responses on the frequency dimension are weighted 0, 1, 2, 3, and 4.

For criterion-referenced scoring on the frequency dimension, the only response considered *unfavorable* is (1) **Have not done.** All other responses are considered *favorable*. A student's score on the scale is the percentage of items to which a favorable response is given.

The responses on the interest dimension are only presented in the form of item frequency distributions.

Career Awareness

MEASUREMENT RATIONALE

In order to understand the opportunities open to him or her in the future, the student must have an awareness of duties, training, abilities and educational requirements of various occupations – along with an understanding of the labor market conditions. Also, skills in gathering such data, such as knowledge of occupational clusters and publications, is necessary for continuing education in this area.

Schools should provide students with guidance to seek out the kind of education and training which will lead them to careers that are best suited to their talents and interests.

Vocational development, for purposes of assessment, is a series of processes involving both the acquisition of knowledge about different kinds of work and the forming of attitudes which will enhance one's chances of succeeding in the work-a-day world.

In the initial stages of vocational maturity students become aware of different kinds of work and workers. This is followed by a growing understanding of the relatedness of educational and occupational opportunities. The more vocationally mature students will actively seek information, accept personal responsibility for career decision and finally base their career choices upon a realistic appraisal of their interests, achievements and aptitudes.

It is assumed that the awareness of jobs and job opportunities, measured at all grades, will lead to the development of vocational maturity.

GENERAL SCALE DESCRIPTION*

These scales contain 40-42 items which focus on the student's knowledge concerning the duties, conditions, training, salary and educational requirements of various occupations. The multiple-choice format requires the student to select the best answer from alternatives. One point is given for each correct response. No adjustment is made for guessing.

---

*These scales were developed by Francis Reardon, Gregory Shannon, Richard Kohr and James Masters from the Department of Education. Some items are used with permission: Copyright, 1976, Scholastic Testing Service, Bensenville, Illinois.

Sample items:

Grade 5

For which of these workers is the ability to spell correctly most important?

A mail carrier
B waitress
C secretary
D bookkeeper

Which job does NOT require a college degree?

A computer operator
B veterinarian
C lawyer
D teacher

Which worker would spend a lot of time working with a sewing machine?

A upholsterer
B clothing salesperson
C laundry worker
D model

Grades 8 and 11

Which of the following requires a college degree?

A bank teller
B bookkeeper
C economist
D transit clerk

Who is most likely to have the highest pay?

A beautician
B physical therapist
C pharmacist
D dentist

The need to work with others as part of a team would be most important to a

A cartoonist.
B movie editor.
C pizza baker.
D fire fighter.

A dietitian would most likely

A plan meals.
B run a weight-reducing studio.
C serve food in a restaurant.
D prepare food.

Appreciating Human Accomplishments

MEASUREMENT RATIONALE

Appreciation of an area of human accomplishment can be expressed in a variety of ways, ranging from that of believing the area to have some minimal value to society, to that of committing a great deal of one's own time and energy to the area. Although it would seem most desirable to define "appreciation" for this goal as depth of commitment to an area, this would present a number of problems. For example, other responsibilities of students outside of school and the geographical location of their homes could affect greatly the opportunities available to them to appreciate accomplishments.

The instrument being used to measure the goal area, however, would be expected to be highly indicative of present or future commitment to areas of accomplishment. The instrument assesses attitudinal openness to human accomplishments. Items contained in it reflect willingness to receive accomplishments, including willingness to do such things as listen to, watch, look at, read or go to significant accomplishments. In addition, items are employed measuring a somewhat higher level of openness, reflecting willingness to learn about accomplishments and what is entailed in producing them.

GENERAL SCALE DESCRIPTION* - Grade 5

The scale contains 57 items which ask pupils whether they would want to receive or learn about human accomplishments of many types. The response options available to pupils are: (1) I *would* want to, (2) I *might* want to, (3) I *don't* know, (4) I *don't think* I'd want to, and (5) I *would not* want to.

The scale is made up of four item groups measuring attitudes toward the accomplishment areas of science (15 items), social science (15 items), fine arts (21 items), and literature (6 items). In addition, the fine arts items can be grouped into the specific areas of music (6 items), art (6 items), drama (3 items), dance (3 items), and film (3 items).

Sample items:

Would you want to:

Read a science magazine?
Watch an artist make an art object or painting?
Go to a play that has a lot of songs in it?

---

*These scales were developed by James Dorwart, James Masters and BarbamDavis from the Division of Research with the aid of many PDE curriculum specialists.

For norm-referenced scoring, the item weighting scheme is:

Response Options

I Would Want To	I Might Want To	I Don't Know	I Don't Think I'd Want To	I Would Not Want To
4	3	2	1	0

For criterion-referenced scoring, the following scoring scheme is used:

Response Options

I Would Want To	I Might Want To	I Don't Know	I Don't Think I'd Want To	I Would Not Want To
1	1	0	0	0

GENERAL SCALE DESCRIPTION* - Grades 8 and 11

The scales contain 64 items measuring student attitudes toward human accomplishments in a number of areas. The response options available to students are: (1) Definitely Would, (2) Probably Would, (3) Uncertain, (4) Probably Would Not, and (5) Definitely Would Not.

The scales are composed of four item groups reflecting attitudes toward the accomplishment areas of science (16 items), social science (16 items), fine arts (24 items) and literature (8 items). The fine arts items can be grouped into the specific areas of music (8 items), art (8 items), drama (3 items), dance (3 items) and film (2 items).

Sample items:

Would you want to:

See a world-famous play?
Talk to officials to learn about your local government?
Read stories written by authors from other parts of the world?
Learn more about what makes a movie great?

For norm-referenced scoring, the item weighting scheme is:

Response Options

Definitely Would	Probably Would	Uncertain	Probably Would Not	Definitely Would Not
4	3	2	1	0

For criterion-referenced scoring, the following scoring scheme is used:

Response Options

Definitely Would	Probably Would	Uncertain	Probably Would Not	Definitely Would Not
1	1	0	0	0

*These scales were developed by James Dorwart, James Masters and BarbaraDavis from the Division of Research with the aid of many PDE curriculum specialists.

Knowledge of Human Accomplishments

MEASUREMENT RATIONALE

A necessary prerequisite to appreciation is the recognition of persons, theories or ideas, works, inventions or phenomena. Students are expected to be able to recognize the names of prominent persons and concepts to which they should have been exposed within the school setting.

It is anticipated that an increased knowledge of the accomplishments of others will enhance the appreciation of their importance to society.

Although some students may have a broad background in one area of human accomplishment – such as art, music, science, drama or government – it is deemed necessary for individuals to have been exposed to a large number of areas of human accomplishment.

To measure the lowest level of knowledge of these accomplishments, students are asked to simply categorize the name, work, theory, idea, invention or phenomena into a general classification indicating that they have had at least a brief exposure. This general awareness is considered a minimal prerequisite for appreciation.

GENERAL SCALE DESCRIPTION*

These scales contain items giving the names of persons, theories, ideas, works, inventions, or phenomena associated with human accomplishment.

There are five general classifications for each of the items. Names which tended to overlap classes were omitted from the list, for example: Benjamin Franklin, Thomas Jefferson and Winston Churchill.

The classifications are:

Grade 5	Grade 8	Grade 11
M = Music or Dance	M = Music or Dance	M = Music or Dance
S = Science or Medicine	S = Science, Math or Medicine	S = Science, Math or Medicine
A = Art	A = Art (Visual)	V = Visual Arts
W = Writing, Books or Plays	L = Literature or Drama	L = Literature or Drama
G = Government or Politics	G = Government or Politics	G = Government or Politics

*This scale was developed by Carl Guerriero, James Hertzog and Kenneth Miller from the Division of Educational Quality Assessment.

Sample items:

Grade 5	Grade 8	Grade 11
John Philip Sousa	Electronics	N.A.T.O.
Electronics	Martin Luther King	Swan Lake
Legislature	Nutcracker Suite	Ralph Bunche
Martin Luther King	Mona Lisa	Louisa May Alcott
Rembrandt	Alex Haley	George Gershwin
Abraham Lincoln	Marie Curie	Norman Rockwell
Edgar Allan Poe	N.A.T.O.	Euclid
Nutcracker Suite	William Shakespeare	Hooke's Law

One point is given for each correct response. No adjustment is made for guessing.

Grade 11 students are also asked to indicate the degree of importance each of the items has in their life, based on the following ranking: (1) Very Important, (2) Somewhat Important, (3) A Little Important, and (4) Not Important.

These data are presented only in the form of response frequency distributions.

Information Usage

MEASUREMENT RATIONALE

In today's world, one skill which is highly important is decision-making ability (analytical thinking). Each time a person comes to a choice between two or more courses of action, he/she is confronted by a problem that must be solved. As the world becomes increasingly complex, individuals are faced with an accelerating number of decisions. Voting for a candidate merely because he or she is a member of a particular political party, eating a certain brand of food because a well-known sports figure endorses it, choosing a career without exploring the opportunities available in it are all examples of poor decision-making. Students must be taught to examine problems in a logical way, to identify them properly, to seek whatever information is needed to establish facts, to be aware of the consequences of decisions and to have the ability to choose courses of action which are most congruent with their own values and desires. Assessment in this area, therefore, focuses upon processes involved in logical decision-making.

GENERAL SCALE DESCRIPTION*

These scales are made up of short, written descriptions of problem situations involving a central character. Students are asked to think about the information presented and to answer from two to four multiple-choice questions under each situation. For certain situations employed at grades 8 and 11, information is presented in tabular form. A total of 33 items make up the grade 5 scale; 32 items are included at grades 8 and 11.

The scales are composed of the following three item groups:

Drawing inferences items measure the ability to distinguish valid from invalid conclusions drawn from the information presented.

Information seeking contains items for which students are asked to study the information presented to them in a situation and to then make judgments about such things as when more information is needed to solve a problem, which information would be most or least useful, and where to obtain the most reliable information.

Consequences and values includes items for which students must determine the most likely consequences of a course of action initiated to solve a problem. In addition, after having been presented with a set of values and desires of the central character, they must use this information to determine which decision would be best for him or her.

---

*These scales were developed by James Masters, James Dorwart and Gregory Shannon from the Division of Research.

One point is given for each correct response. No adjustment is made for guessing.

Sample situation:

John will soon graduate from high school. He is thinking about going to college. His parents do not have enough money to pay for this. His uncle has offered him a job which pays well. John knows he would enjoy this job because he would learn a great deal about the kinds of things he would study if he went to college. But, John still wants to go to some kind of school. He reads an advertisement in a magazine about a training program offered by the ABC Training School. This school says that it has trained thousands of people for high-paying careers. The cost of their training program is $2,000.

1. **Which of the following are you MOST sure about?**

   A. If John takes ABC's training program, he will make a lot of money.
   B. ABC's training program is a good one since it costs so much money.
   C. John believes that education is important.
   D. John's parents want him to go to college.

2. **Which of the following is likely to be of the LEAST use to John in finding out how good the ABC training program is?**

   A. Read more magazine advertisements about the school.
   B. Contact the Better Business Bureau office nearest the school.
   C. Talk to students who have graduated from the training program.
   D. Talk to possible employers of the school's graduates.

3. **John believes that education is most valuable when the things you are learning can be used at the same time you are learning them. He does feel it is important for him to become trained for a career, but he does NOT like the idea of being in debt. What should he do?**

   A. Go to college part-time while earning money washing dishes in a restaurant.
   B. Work for his uncle part of the year to save money to go to college the rest of the year.
   C. Borrow money and take ABC's training program.
   D. Borrow money to go to college.

Section Two: Measurement Safeguards

DESIGNING AN ASSESSMENT PACKAGE

OVERALL PICTURE

Each year during the first two weeks of March, about 120,000 students in about 1,000 Commonwealth schools complete the EQA inventory. Do the outcomes summarized in the *Manual for Interpreting School Reports* accurately reflect student progress? To answer this, one must know the answers to questions such as: Who provided recommendations for test content? What do they really measure and how accurately do they measure it? How does matrix sampling work? The following sections highlight the safeguards used to produce high quality instruments and how the instruments stood up to these checks.

A WORD ABOUT ATTITUDES

Attitudes, beliefs, values, etc., are abstractions. Nevertheless, they are real enough to each individual holding them. They are typically thought of as a state of readiness–a predisposition to act or react in a certain way when faced with certain situations. A person's attitudes are always present but remain dormant most of the time. They are expressed in speech or other behavior only when the object of the attitude is perceived. A person may have strong attitudes for or against astrology but actively express them only when some issue connected with astrology arises–or when confronted by an attitude scale! Attitudes are often reinforced by beliefs (the cognitive component) and attract strong feelings (the emotional component) that will lead to particular behaviors (the action tendency component).

The measurement of attitudes always involves making inferences. Since the attitudes cannot be seen or measured directly, we must infer their presence from consistencies that appear in the individual's behavior. Observing individuals across time in everyday situations is probably the best way to learn how the individual thinks, feels and acts. Clearly this method is much too cumbersome and costly when we want to investigate the intensity and direction of attitudes on a large number of people, forcing us to rely instead on pencil-paper measures of the individuals concerned.

WHAT ABOUT PAPER-AND-PENCIL MEASURES OF ATTITUDES?

The use of paper-and-pencil techniques for measuring attitudes is often questioned. These questions are directed at both the measuring instrument and the students responding to them. Critics are concerned with the possibilities that:

1. People misunderstand what the items are asking.
2. People don't always tell the truth on this type of measurement.
3. Scores on these instruments are seldom presented in any meaningful way.
4. People might not respond consistently to similar items.
5. People might respond to items differently at various times.

In the development of the Educational Quality Assessment Inventory (EQAI) scales, these and other concerns were taken into consideration.

From the outset, all pilot instruments were put through an obstacle course of checks and balances designed to determine their susceptibility to various errors of measurement. The test instruments were then revised and submitted for additional field trials. This philosophy of *testing the tests* before using them required strong cooperation between the Department of Education and over 40 Commonwealth school districts before the present test package (the third version since EQA testing began in 1969) was ready for use in actual assessment.

YES, BUT CAN THEY READ IT?

If students cannot read the test, the tests cannot assess the students. Although it is impossible to control the range of verbal comprehension in a program assessing over 40,000 students per grade level each year, it is feasible to develop scales that can be read and understood by the vast majority of respondents.

Toward this end, monitors conducting field trials were required to submit detailed reports on the understandability of directions, item sentence structure and vocabulary for each instrument. Also, students were asked to comment on the scales and identify words or items that didn't make sense to them. Changes resulting from these data increased the readability of all EQA instruments. Although the reading level indices were not computed on the present EQA test package, it should be noted that Gunning-Fog indices computed on the 1974-1977 test package showed that seven scales within the Grade 5 inventory had grade level values ranging from 1.8 through 4.0. At Grade 8, Gunning-Fog readability indices ranged from 4.0 through 5.0 for nine of the scales with two instruments having readabilities of 6.0 and 6.7. In the Grade 11 inventory, seven scales had readability indices ranging from 4.0 through 5.0 with three instruments requiring a level range from 6.0 through 7.2. Results from a study involving a limited number of special education students in a large suburban school district indicate that the majority of the tests comprising the previous battery would need additional revision before becoming appropriate for our special education population. Because of the attention given to reading level during the construction and field testing of the revised EQAI, it is probable that the values cited above are representative.

FROM ITEM CONTENT TO ANSWER SHEET

Handing someone a list of attitude questions, waiting 10 minutes, then collecting it does not assure that the marks in the answer column were made solely in response to the contents of statements on the list. Those other factors influencing answers are called response sets. A common response set is the tendency of respondents to tell you what they think you want to hear (make a good impression). In their development stages, the scales and items used from 1974-1977 were checked against a special 36-item instrument designed to pick up the tendency to give socially desirable answers. This scale contained such questions as:

1. I never forget to say please and thank you.
2. Sometimes I don't like to obey my parents.
3. I always finish my homework on time.
4. Sometimes I do things I have been told not to do.

Those items and scales whose scores are found to be associated with socially desirable responses were deleted or revised to minimize the relationships found. The correlation coefficients between the final total instrument scores and a 30-item version of the social desirability scale ranged from .11 through .19 for Grade 11 and from .04 through .08 for Grade 8. Although not used for the present test package, it would be reasonable to expect that socially desirable responding by students is associated with responses on the current attitude scales at about the same level as found previously. In fact, the matrix sampling process which reduces the number of items each student must respond to probably serves to minimize this type of problem.

An attempt was made to minimize the effect of position bias by including both positively and negatively worded items in the goal scales. The appreciating human accomplishments scale and creative activities checklists are the only EQAI instruments which do not contain reflected items. An estimate of response position bias can be determined by counting the number of *perfect scores*. To get a perfect score, the student must answer in the most positive direction for each of the separate items–a highly questionable feat. Fewer than one percent of the nearly 40,000 respondents obtained the highest possible score on the form of the test which they took.

FROM ITEM ANSWERS TO SCORES

To give school personnel a clearer picture about the performance of their students on the EQAI, two scoring methods are used. In the first scoring method cognitive instruments are scored according to the simple number correct. For attitude scales, this method organizes the response options into a continuum. Different scoring weights are then applied to each point on the continuum. Consequently, for an item such as "I like school" with the response options (a) Strongly Agree, (b) Mostly Agree, (c) Mostly Disagree and (d) Strongly Disagree, weights of 3, 2, 1 and 0 are applied respectively to each answer choice. This method is based on the assumption that strongly agreeing with the statement is more positive than mostly agreeing with the statement.

Item scores obtained by this method are summed and used to give norm-referenced information about student performance. How well a group of students performs on the scale is determined by the relationship of their estimated mean scores to the estimated mean scores achieved by other groups of students. This *norm-performance* tells very little about favorableness of student responses.

To obtain information relative to a criterion, one can apply a different scoring scheme. For attitudinal measures, this scoring scheme is based on the notion that each item within the scale offers the respondent the chance to show a positive or negative attitude toward the specific content presented by the item. Hence, the response choices to the above attitudinal item are scored by assigning a "1" to both the strongly agree and mostly agree choices and a "0" to the remaining choices. For cognitive measures, a positive weight (a "1") is applied to the correct response. The number of positive responses given by each student is compared to a criterion independently determined. If the number of favorable responses (in the case of an attitudinal measure) or the number of correct responses (as in a cognitive measure) meets or exceeds the criterion, the student is said to have achieved at criterion. In the case of scales used in the EQAI, three criteria were set. Level 1 requires students to respond correctly or favorably to one-third or more of the items in the scale; Level 2 requires correct or favorable responses to more than half of the items; and Level 3 requires correct or favorable responses to two-thirds or more of the items.

REVIEW OF 1974-1977 PACKAGE

To prepare for the large-scale revisions to the EQAI package, EQA staff assembled information based on their collective experience with school personnel and on surveys which were follow-ups to each year's assessment. The concerns and desires that became evident were incorporated into the development stages of the present instrument package. Item analyses were also conducted on each instrument used during 1974-1977 to examine items with respect to the percentage responding to each choice, the item to total correlations, cultural bias, etc. Items identified as statistically weak became prime candidates for removal or alteration.

REVIEW COMMITTEES

To provide guidance in the development of the present measurement package, various review committees were formed. The composition of these committees varied according to the area and the extent to which an instrument was being revised. For example, reading, writing skills, and mathematics, where three totally new tests were being prepared, the committees had greater involvement, since they were responsible for recommending objectives as well as for reviewing test items. Although the tasks engaged in by particular committees differed, the general review included reacting to the EQA instrument in use from 1974 to 1977, offering suggestions on content not covered, determining objectives to be measured, evaluating prospective items and, in some instances, assisting in the item writing. For example, in mathematics, the review committee was supplied an extensive list of objectives based on the department's Pennsylvania Retrieval of Information for Mathematics Education (PRIMES) project. They reacted to each items of degree of importance, whether the objective should be measured at the conceptual, computational or application levels and grade-level appropriateness.

During a meeting at Harrisburg, each objective was discussed, the ratings were shared and discrepancies were resolved. The 60 most important objectives were selected, and at least two items were written per objective for each grade level. These items were submitted to a reviewer representing each grade level. The items were further refined on the basis of reviewer's comments. The revised items were then submitted to yet another individual, who reviewed the items for all three grade levels, not only for item adequacy but for whether the items seemed to be at an appropriate difficulty level across grades. The items were again revised before being deemed ready for a field test.

A brief summary of the groups involved in the review follows, and the individuals who comprised these groups are identified in Appendix A. The review committees provided an extremely valuable service in guiding EQAI construction efforts. However, responsibility for the final instrumentation belongs to EQA staff.

Area	*Types of Individuals Involved in Review Process*
Self-Esteem	Secondary and elementary teachers who responded to an invitation to review items.
Understanding Others	Office of Equal Rights

Reading	Reading specialists at the university, intermediate unit and public school levels and the State Advisory Committee on Reading.
Writing Skills	Individuals who were language arts specialists at the university, intermediate unit and public school levels.
Mathematics	Individuals who were elementary and secondary mathematics specialists at the university, intermediate unit and public school levels.
Interest in School and Learning	Secondary and elementary teachers who responded to an invitation to review items.
Societal Responsibility and Knowledge of Law/Government	Individuals who were social studies specialists at the university, intermediate unit and public school levels.
Health	Individuals from the Pennsylvania State Association of Health, Physical Education and Recreation.
Creative Activities	Fine arts advisors from PDE, Bureau of Curriculum Services.
Career Awareness	Individuals broadly representative of career development areas at the university, intermediate unit, public school and vocational-technical levels.
Appreciating Human Accomplishments and Knowledge of Human Accomplishments	Curriculum specialists in PDE's Bureau of Curriculum Services.
Information Usage	Division of Research

CONTENT REVISIONS

From recommendations by special advisory committees and people in Commonwealth school districts, a plan for changes in test content was developed. The following describe these changes:

Self-Esteem–The response from people in the field who used the 1974-1977 instrument package indicated that items not having a school context provided little useful information. Thus, an item like "Things are all mixed up in my life," which could relate to nearly anything in the student's life, was not viewed as particularly affectable by the school situation. An item like "My

teachers make me feel that my ideas are worthwhile" does provide a school context and yields useful information, in that school people feel that this is an area that the school can influence. Therefore, the changes reflected a shift in item content toward that of a school context.

Understanding Others–EQA's experience with this area during 1974-1977 showed that the content was too sensitive with respect to the handicap items. Therefore, the decision was made to remove the handicap items and replace them with items on physical characteristics. Field-testing also showed that the loss of sensitive items in racial and religious areas, while desirable from one standpoint, leads to a rather innocuous set of items with marginal discriminatory power. In addition, items pertaining to differences in intellectual status were replaced by items concerning cultural tradition.

Reading, Writing Skills, and Mathematics–Because of the increased interest in cognitive skills, an enlarged test for both the verbal and math areas as well as a new test in written expression was needed. Experience in the field clearly showed that, while correlated with standardized tests, verbal analogies provide little useful information to school people. Therefore, the new verbal instrument measures reading comprehension. The math instrument has content measuring conceptual, computational and application areas and consists of 60 items rather than the previous 30 items. An effort was made to design foils which reflect common, incorrect strategies for arriving at a math answer. The writing skills instrument deals with higher-level cognitive processes that differ somewhat from the typical language arts test but tap the prerequisite skills deemed necessary for coherent written material.

Interest in School and Learning--As with Self-Esteem, items not having a school context provided little useful information to teachers and administrators. This is particularly true of items dealing with a general attitude toward learning. To enhance the sense of "ownership," new items having a definite school context were developed.

Societal Responsibility and Knowledge of Law/Government–Although EQA had several meetings with the Pennsylvania Council for Social Studies, little concensus as to the apropriate content emerged. Some people favored keeping the Societal Responsibility instrument but making some content revisions. Some felt that a cognitive instrument should be developed. Most were reluctant to completely remove an attitudinal measure. The issues were resolved by making some content and response choice changes in the attitudinal scale and introducing a new cognitive measure which covers government, law, rights and responsibilities, and interdependence.

Health–At grades 8 and 11, the changes include the introduction of items on physical activity and an alteration to some of the items to represent a general improvement in content. At grade 5, the health knowledge test was expanded to reflect an increase in the content coverage.

Creative Activities–The changes were primarily improved content balance and an improvement in the response choices offered to students.

Career Awareness–At all three grade levels, the test instrument measures the knowledge base used in making career choices. The content was expanded over past instruments.

Appreciation and Knowledge of Human Accomplishments–The changes include the introduction of a knowledge instrument as well as improved content coverage in the attitudinal measure used from 1974-1977.

Information Usage–In view of the general concern over cognitive development and interest in the measurement of analytical thinking, a measure of information usage was judged to be appropriate. Thus, the instrument was designed to measure the ability of the student to make inferences from given facts, gather appropriate information, and understand consequences of decisions.

ITEM WRITING

Proposed items for instruments to be revised or developed came from many people. Teams composed of staff from EQA, the Division of Research and, for particular areas, the Division of Curriculum Services were charged with the responsibility of working with the review committees in writing and evaluating items. Certain general guidelines were followed in writing items, regardless of the area. The vocabulary and language structure used in the items had to be (1) appropriate to the grade level, and for grade 5, designed for a fourth grade or lower reading level, (2) free of sexism in the language used or in the content of the questions, (3) free from cultural and geographic bias and (4) grammatically correct.

FIELD TESTING/STUDENT INTERVIEWS

Over 40 school districts volunteered their students for field-testing. The number of students responding to a given field test instrument ranged from 200 to 400, representing rural, urban and suburban areas. During field testing, the students were carefully observed by monitors, who were classroom teachers or PDE staff. The amount of time required for students to complete the instrument was recorded, along with student behavior rated by the following dimensions:

1. Bored—Interested
2. Careless—Careful
3. Distracted—Concentrating
4. Confused—Clear
5. Did not take test seriously—Took test seriously

The group of students tested was also rated as to whether its reading ability could be characterised as above average, below average, well below average, or wide mixture of levels. The monitors provided other written comments concerning special situations or problems encountered with the testing situation. Teachers who served as monitors were also asked to look over the questions and comment on (a) appropriateness, (b) understandability and (c) vocabulary level.

To supplement the information derived from having students respond to the instrument, two additional activities were devised. One was a special written task in which four to six students per classroom were asked to read a special subset of the questions and report on a special answer sheet (a) whether they understood the questions, (b) their answer to the question, and (c) a brief explanation of why they answered a particular question the way they did. They were also asked to identify any words they did not understand.

The second activity involved student interviews. Generally, six to eight students were interviewed by PDE staff following the testing. Questions were posed as to (a) what the student thought about the questionnaire (b) whether there were certain questions that were confusing, (c) how

certain questions were answered and why they were answered in a particular way, (d) the relevance of content embedded in the items, and (e) specific information that was tailor-made to the particular area tapped by the instrument. For example, in interest in school and learning, questions regarding what is especially liked or disliked about school, relationships with teachers, the kind of student activities in which one is involved were posed.

Information from the written response task and student interviews was especially valuable in identifying problem items or other deficiencies. Changes in content were made to resolve any problems noted.

ITEM ANALYSIS

Following the field test, the answer sheets were scanned and student item responses were recorded on computer tape. Statistical analyses were then conducted with particular attention given to (a) the percentage getting an item correct for cognitive tests or the item mean for non-cognitive scales, (b) item-total correlation, and (c) item-subscale correlation, where a subscale is a score obtained by summing across a subset of items which pertain to a particular subdivision of the test. Analyses also compared the statistical results for groups based on sex, type of community and reading level. These data were taken into account when recommending that items be accepted, revised or rejected.

FINAL SELECTION OF ITEMS

A final selection of items was made by combining the information gained from teacher comments, student interviews and special written tasks and statistical analysis of field test data. A special effort was made to preserve the content integrity of each instrument so that the items comprising the final version represent the concepts, skills, or attitudinal areas specified in the original design.

Because of the matrix sampling procedure, there was yet another step in the process–allocating items to test forms. This process used item mean or precent-correct data in conjunction with content considerations to insure that the forms be as balanced as possible. See the next section for a fuller discussion of the matrix sampling process.

MATRIX SAMPLING TEST DESIGN

This section of the manual describes the matrix sampling procedure, the rationale for using matrix sampling, the process of assigning items to forms, the equivalence of forms, and the method for estimating school means scores.

MATRIX SAMPLING PROCEDURE

From 1969 to 1977, the EQA process was based on "census" testing in which all students in the grade level being tested took *all* the items in the test package. In matrix sampling, each student takes only a portion of the total number of items in the scale. This is accomplished by taking the complete set of items for a given goal area and assigning an equal number of items to separate forms. Thus, in grade 5, where there are three forms, 20 of the 60 math items were assigned to each form. During administration, the questionnaire booklets are packaged so that every third 5th grader would receive a different form. In a given elementary school, about one-third of the 5th grade students receive Form A, another third receive Form B, and a final third Form C. Since all forms are identical in format, test administration proceeds in the same straightforward manner as in census testing.

To further explain the meaning of the word "matrix, " several figures have been prepared. A matrix is a two-dimensional array with one dimension relating to test items (columns) and the other (rows) to students. Figure 1 shows a matrix having arrays consisting of N students and K items. In conventional census testing, *all* students would have taken all items in the test. Figure 2 shows the type of achievement data gathered by census testing for a hypothetical nine-item test. A '1' represents a correct response and a '0' denotes an incorrect response.

Figure 3 shows data resulting from a matrix sampling test procedure. In this illustration there are three items for each of three forms. The figure shows that one-third of the students received items in each of the three forms. Thus, all students are tested (on one form or another) and all items are administered (to a portion of the students).

RATIONALE FOR MATRIX SAMPLING

Matrix sampling has a number of advantages over census testing. Since each student takes only a portion of the items, testing time is shorter. Currently in EQA, with four forms at grades 8 and 11, testing time is about two hours, and at grade 5, with three forms, testing requires about two and one-half hours, substantially under the four to four and one-half hours required previously.

Shortening the test for each student has some definite advantages. Fatigue effects, built up from taking many items of similar type (for a particular area) and from the sheer length of a large test package, are greatly reduced. A greater time-to-item ratio removes the tendency for some tests to be speeded. Both of these effects enhance validity. Furthermore, a shorter testing time does not intrude as much on the school's regular activities. For example, at grades 8 and 11 only one day is required for testing, rather than several days.

Another strength of matrix sampling is better content coverage. An expansion of the number of items means a greater range of content. For example, the EQA math test was expanded from 30 to 60 items at each grade level; this permitted EQA to design a measure with a much broader range of skills.

ASSIGNMENT OF ITEMS TO FORMS

All forms were prepared with the same format to facilitate administration. Each form has the same number of items for a particular goal area. When items are assigned to forms, the following considerations are observed:

1. Each form should have approximately the same number of positively and negatively stated items in the case of noncognitive instruments. For cognitive tests, each form should have a balance of keyed responses and should be different from form to form.

2. Each form should be similar in content structure. For example, in the Self-Esteem area, each form would have approximately the same number of items depicting feelings of acceptance and about the same number tapping feelings of success. In a cognitive area, each form would have about the same number of items per content area.

3. Each form should be similar in terms of difficulty level (average item mean in the case of noncognitive instruments). Field test data were used as a basis for this consideration in item assignment.

FORM EQUIVALENCE

Although attempts are made when designing a test to have form equivalence, it is necessary to measure a large sample of students in order to determine whether the forms are similar statistically. During the 1978 assessment approximately 30,000 students were tested at each grade level. Therefore, approximately 7,500 eighth and eleventh graders responded to each of four forms and about 10,000 fifth graders to each of three forms. From a sample of about 3,500 students taking each form, statistics were compiled which compare the forms with respect to mean score, variability and internal consistency. These results, shown for each grade level in Tables 1, 2 and 3, indicate that overall there is very close comparability of forms. This may be observed through the very close values in the means, standard deviations and reliability coefficients. Differences between form means were evaluated by computing an effect-size index rather than employing a significance test since virtually all differences can be found to be statistically significant when samples are extremely large. The index for effect size (Cohen, 1969) is a ratio found by dividing the observed difference between means by the common standard deviation. The differences are thus converted into standard deviation units. Differences of a quarter of a standard deviation (0.25 or less) were regarded as being of minimal consequence. At grade 5, with three forms, only 6 of the 42 possible comparisons exceeded the .25 figure; most values ran from .02 to .15.

The largest deviation occurred for Creative Activities, where Forms B and C differed by .77 of a standard deviation. At grade 8, with four forms, 20 of the 84 possible combinations exceeded the .25 figure; about half of the comparisons ranged from .00 to .15. The largest deviation occurred for Health and Safety Practices, where the difference was .64 of a standard deviation. At grade 11, with four forms, 16 of the 84 possible combinations exceeded the .25 figure; over half of the indexes ranged from .01 to .15. The largest deviation occurred for Career Awareness, where the difference was .66 of a standard deviation.

ESTIMATING SCHOOL MEANS

On the EQA School Report, a score for a school is expressed as a mean or average student score. This value is an estimate of what school mean would be if all students had taken all the items. An example is shown in Figure 4, which displays a matrix of data gathered from 10 students, four of whom took Form A, three who received Form B and three who responded to Form C. Each form consists of three different test items. O's and 1's refer to incorrect and correct answers respectively. Notice that an additional column, labeled TOT, for each form refers to a total score produced by adding the scores on the individual items. Student one achieved a score of 2 because items 1 and 3 were correct. Student six, who responded to Form B, has a score of 0 because none of the answers was correct. To calculate an estimated school mean, use the following steps:

1. Determine a form score for each student.
2. Determine the mean student score for each form.
3. Produce a weighted estimate of a school mean based on the following formula:

$$\hat{\bar{X}} = \frac{\sum_{i=1}^{F} (N_i \bar{X}_i) F}{\sum_{i=1}^{F} N_i}$$

Where,

N_i refers to the number of students responding to a given form.

$\bar{X}_i$ refers to the mean score achieved by students responding to a given form.

F refers to the number of forms.

Σ refers to operation of summation.

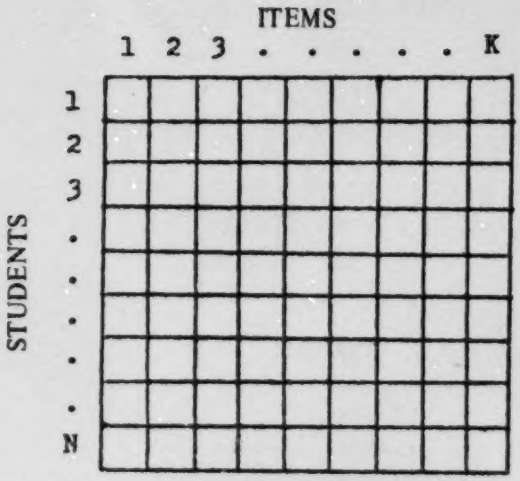


Figure 1. A matrix showing a population of N students (rows) and a universe of K items (columns)

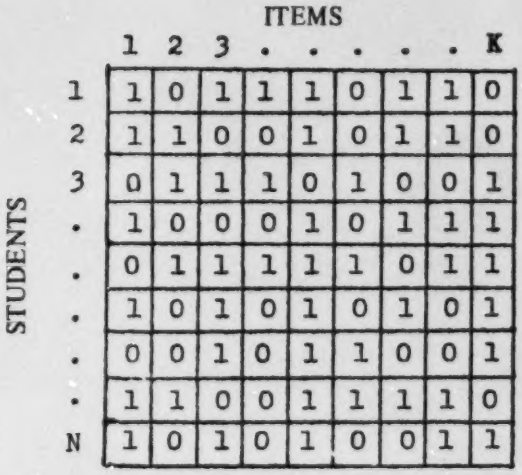


Figure 2. A typical data matrix obtained through census testing, when all students take all items.

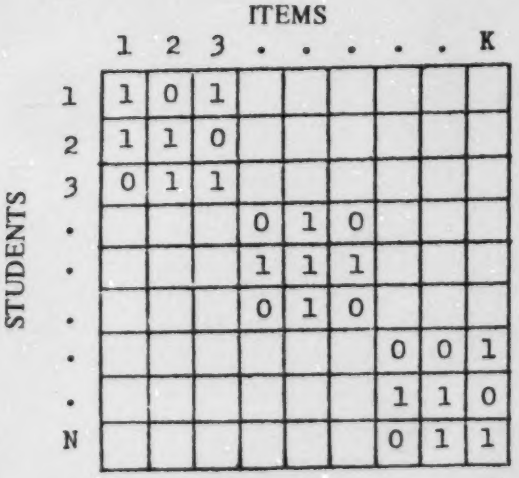


Figure 3. A typical data matrix obtained from a matrix sampling test administration. All students are involved in the testing and all items are represented in the assessment.

Figure 4

Calculation of Estimated School Mean Based on Multiple Matrix Sampling Procedure

		Form A				Form B				Form C			
		1	2	3	TOT	4	5	6	TOT	7	8	9	TOT
Students Taking Form A	1	1	0	1	2								
	2	0	1	1	2								
	3	0	1	0	1								
	4	1	1	1	3								
Students Taking Form B	5					1	1	1	3				
	6					0	0	0	0				
	7					1	0	1	2				
Students Taking Form C	8									1	0	1	2
	9									1	1	1	3
	10									1	1	0	2

Mean score for students taking Form A = 2 + 2 + 1 + 3 = 8/4 = 2.00

Mean score for students taking Form B = 3 + 0 + 2 = 5/3 = 1.67

Mean score for students taking Form C = 3 + 2 + 2 = 7/3 = 2.33

Estimated School Mean $(\hat{X})$:

$$\hat{X} = \frac{[4(2.0) + 3(1.67) + 3(2.33)]\ 3}{4 + 3 + 3}$$

$$\hat{X} = \frac{[8 + 5 + 7]\ 3}{10} = \frac{(20)\ 3}{10} = \frac{60}{10}$$

$$\hat{X} = 6.0$$

Table 1

Comparability of Test Forms: Grade 5

Area	Items/ Form	Form A			Form B			Form C		
		Mean	SD	Rel	Mean	SD	Rel	Mean	SD	Rel
Self-Esteem	11	21.04	4.85	.61	20.80	4.88	.63	20.24	5.00	.62
Understanding Others	15	40.61	11.46	.82	40.91	11.70	.81	37.42	10.06	.73
Reading	16	8.65	3.88	.80	8.97	3.68	.78	9.42	3.81	.80
Writing Skills	15	9.52	3.16	.72	9.40	3.11	.71	9.71	3.06	.71
Mathematics	20	12.40	3.60	.74	12.20	3.45	.72	11.65	3.50	.72
Interest in School and Learning	9	18.57	4.18	.71	18.67	3.88	.61	18.17	4.07	.67
Societal Responsibility	9	14.60	3.09	.72	14.13	3.13	.70	14.07	3.40	.76
Knowledge of Law/ Government	7	3.40	1.53	.36	3.60	1.71	.51	3.75	1.71	.52
Health	15	8.46	2.74	.61	10.02	2.54	.58	10.46	2.66	.65
Creative Activities	12	19.04	8.33	.73	20.29	8.03	.74	14.18	7.83	.77
Career Awareness	14	7.97	2.77	.64	8.33	2.65	.59	8.07	2.54	.59
Appreciating Human Accomplishments	19	48.46	13.30	.84	49.05	13.99	.85	50.14	14.04	.87
Knowledge of Human Accomplishments	15	7.46	3.14	.74	6.80	3.04	.68	7.58	2.98	.70
Information Usage	11	6.34	2.68	.70	6.04	2.61	.68	6.10	2.66	.70

No. students in each item analysis sample: Form A (N=3017), Form B (N=3004), Form C (N=2964)

Table 2

Comparability of Test Forms: Grade 8

Area	Items/ Form	Form A			Form B			Form C			Form D		
		Mean	SD	Rel	Mean	SD	Rel	Mean	SD	Rel	Mean	SD	Rel
Self-Esteem	8	14.44	3.39	.56	14.53	3.64	.59	14.81	3.58	.61	14.48	3.86	.66
Understanding Others	10	27.65	7.88	.80	26.36	8.28	.80	29.84	7.55	.81	28.64	8.41	.83
Reading	12	7.23	3.14	.78	6.60	3.04	.75	6.50	2.92	.72	6.60	2.91	.74
Writing Skills	15	8.77	3.19	.72	9.02	2.94	.65	9.17	3.09	.72	9.36	3.02	.69
Mathematics	15	7.86	2.91	.68	8.34	3.37	.75	7.40	3.26	.74	8.04	3.28	.73
Interest in School and Learning	10	16.64	5.12	.76	16.46	5.27	.77	17.41	4.63	.68	17.10	4.75	.74
Societal Responsibility	12	16.09	4.58	.73	13.63	4.82	.73	15.72	4.44	.70	14.43	4.70	.74
Knowledge of Law/ Government	11	6.30	2.51	.68	5.88	2.50	.66	6.48	2.67	.70	6.21	2.69	.73
Health	12	21.29	5.80	.69	23.46	5.36	.70	19.87	5.79	.72	22.75	5.93	.73
Creative Activities	10	13.39	7.08	.73	10.98	6.76	.75	10.63	7.08	.76	12.14	6.97	.73
Career Awareness	10	5.16	2.03	.47	6.08	2.14	.57	6.12	2.12	.56	5.63	2.15	.56
Appreciating Human Accomplishments	16	33.67	11.38	.86	33.94	10.63	.83	32.33	11.35	.85	31.02	11.03	.85
Knowledge of Human Accomplishments	15	6.95	3.29	.76	7.93	3.16	.76	7.66	3.49	.78	7.75	3.69	.80
Information Usage	8	3.39	1.85	.51	3.54	2.02	.62	3.89	1.88	.52	4.10	1.87	.56

No. students in each item analysis sample: Form A (N=3908), Form B (N=3933), Form C (N=3808), Form D (N=3789)

Table 3

Comparability of Test Forms: Grade 11

Area	Items/ Form	Form A			Form B			Form C			Form D		
		Mean	SD	Rel	Mean	SD	Rel	Mean	SD	Rel	Mean	SD	Rel
Self-Esteem	8	15.25	3.66	.72	14.61	3.76	.70	14.38	3.27	.58	14.68	3.24	.56
Understanding Others	10	28.12	8.75	.84	27.43	8.01	.82	29.46	8.11	.81	29.37	7.90	.83
Reading	12	6.26	2.81	.70	6.50	2.90	.72	6.68	3.11	.77	5.67	2.93	.72
Writing Skills	15	8.28	2.94	.64	8.84	3.14	.71	8.47	3.10	.69	9.15	3.24	.72
Mathematics	15	8.96	2.91	.69	9.15	3.06	.70	8.82	2.91	.66	8.42	3.14	.72
Interest in School and Learning	10	15.56	4.63	.73	16.40	4.64	.68	15.15	5.12	.77	16.37	4.54	.70
Societal Responsibility	9	12.81	3.26	.66	12.03	3.74	.76	12.99	3.37	.68	12.86	3.99	.77
Knowledge of Law/ Government	12	6.23	2.40	.61	6.33	2.90	.74	5.72	2.80	.71	6.49	2.62	.67
Health	12	19.14	6.63	.78	21.09	6.64	.82	20.16	5.34	.70	20.51	5.78	.74
Creative Activities	10	12.46	6.80	.70	10.83	6.69	.74	10.73	6.42	.71	9.21	6.71	.74
Career Awareness	10	6.14	1.92	.48	5.02	1.91	.43	5.34	2.13	.57	6.43	2.33	.65
Appreciating Human Accomplishments	16	32.56	11.16	.86	33.53	11.85	.88	32.73	11.07	.86	33.11	11.20	.86
Knowledge of Human Accomplishments	15	7.03	3.05	.70	6.81	2.95	.70	6.73	3.10	.72	7.59	3.38	.76
Information Usage	8	4.44	1.96	.59	4.33	2.06	.64	4.68	2.17	.70	4.42	2.02	.64

No. students in each item analysis sample: Form A (N=3610), Form B (N=3615), Form C (N=3580), Form D (N=3521)

RELIABILITY

This chapter deals with various indices which provide evidence of reliability. Reliability is that characteristic of a measuring instrument which deals with consistency of scores, such as within the scale itself (internal consistency) or over time (stability). Reliability coefficients are reported as two-place decimal figures ranging from .00 to 1.00. As the instrument increases in reliability, the coefficient increases in value. Generally, reliability coefficients are interpreted as the proportion of the variance in a set of scores which is caused by variation in the examinees' true scores rather than by errors of measurement. The coefficients are derived by taking into account the length of the test and the extent to which test items contribute mutually confirming, or consistent, information. It should be noted that the most meaningful reliability estimate pertains to the manner in which test scores are used. Traditionally, standardized tests are used for making decisions about individual students. Thus, the commonly used procedures for estimating reliaibility which define true scores in terms of students are appropriate. However, it is also possible to define true score variance in terms of other units, such as the school. Since the EQA report is designed for school buildings, reliability estimates which take school differences into account are addressed in this chapter. Another type of information which is included in this chapter and which is related to the issue of reliability is that of standard error. As applied to schools (the unit of decision), the statistic of concern is the standard error of the mean, which permits one to construct with a given probability the limits within which the true score of a school will fall. Finally this chapter provides information regarding the percentage of omitted items for each measuring instrument.

INTERNAL CONSISTENCY

Although the internal consistency estimate of reliability is based on students as the units of analysis, it is presented here because of its wide-spread use and familiarity by most practitioners. The KR-20 and the coefficient alpha reliability formulas could not be used with the present EQA test package because of the matrix sampling process in which individual students only respond to a portion of the total number of items in a particular area. Used instead was a procedure which is outlined by Shoemaker (1973) and which estimates the average reliability of a single item within a form and then inserts this value in the well-known Spearman-Brown formula to obtain the relability of the scores based on the full set of items comprising the measurement of a given goal. Reliability estimates produced for each goal within each school are provided on pages 9-23 of the EQA school report. In Table 4 are the median internal consistency estimates obtained during the 1978 assessment. These values seem quite satisfactory. Of the 42 reliabilities, 21 are in the .90's, 19 are in the .80's and only 2 in the .70's. There are 34 reliabilities in excess of .85.

STABILITY OF SCHOOL MEANS

An approach to reliability which took into account schools as the units of analysis rather than students was the stability of group (school) means. This was done by first dividing the student data into random halves and producing, via the matrix sampling formulas, a mean score for each half. Correlation coefficients between the two mean scores were then calculated. These correlations are shown in Table 5. Coefficients of this type are typically lower than internal consistency estimates. Still, the magnitude of these stability coefficients indicate acceptable reliability.

STANDARD ERROR OF THE MEAN

Since matrix sampling methodology is used in estimations of school mean scores, an appropriate statistical technique is used to develop a standard error of the mean. The magnitude of the standard error of the mean for a particular matrix sampling design will decrease as the number of students increases. Thus, a larger standard error of the mean will occur for smaller schools. Adding and subtracting the standard error of the mean will produce a range within which the true score of a school could be expected to fall with a probability of .68. Table 6 presents the median values for the standard error of the mean achieved by schools in the 1978 assessment cycle.

PERCENTAGE OF OMITS

Although the amount of omitted items rarely occurs in discussions of reliability, it is included here because of the concern expressed by some school people as to whether students were responding to most of the items and how this influenced a school's score. On pages 9-23 of each EQA school building report is the percentage of items omitted for each goal area. The median values are calculated for those schools participating in the 1978 assessment and are presented in Table 7.

As shown, the percentage of omits is less than one percent for all but six of the indices. Note particularly the high figure for Knowledge of Human Accomplishments. This high percentage resulted from counting the *I don't know* response as an omit. Overall, the percentage of items which are omitted is extremely small and has little impact on the reliability of scores.

Table 4

Median Internal Consistency Reliability Estimates-1978

Area	Grade 5	Grade 8	Grade 11
Self-Esteem	.82	.86	.88
Understanding Others	.92	.94	.95
Reading	.91	.92	.90
Writing Skills	.86	.89	.89
Mathematics	.86	.91	.89
Interest in School and Learning	.84	.91	.91
Societal Responsibility	.87	.91	.91
Knowledge of Law/Government	.73	.89	.88
Health	.79	.91	.92
Creative Activities	.89	.91	.91
Career Awareness	.81	.80	.80
Appreciating Human Accomplishments	.94	.95	.96
Knowledge of Human Accomplishments	.91	.95	.96
Information Usage	.86	.82	.87

Table 5

Stability[1] of School Means-1978

Area	Grade 5 (N=457)	Grade 8 (N=151)	Grade 11 (N=128)
Self-Esteem	.48	.52	.52
Understanding Others	.55	.66	.70
Reading	.69	.79	.90
Writing Skills	.68	.79	.87
Mathematics	.76	.89	.83
Interest in School and Learning	.61	.78	.81
Societal Responsibility	.63	.70	.74
Knowledge of Law/Government	.60	.85	.83
Health	.71	.60	.50
Creative Activities	.46	.69	.65
Career Awareness	.66	.81	.82
Appreciating Human Accomplishments	.50	.71	.50
Knowledge of Human Accomplishments	.76	.87	.79
Information Usage	.66	.79	.87

[1]Correlation coefficients between the means of random halves

Table 6

Median Values for the
Standard Error of the Mean-1978

Area	Grade 5 (N=457)	Grade 8 (N=151)	Grade 11 (N=128)
Self-Esteem	0.74	0.69	0.62
Understanding Others	1.08	0.97	0.93
Reading	0.36	0.38	0.38
Writing Skills	0.35	0.38	0.39
Mathematics	0.33	0.37	0.37
Interest in School and Learning	0.62	0.68	0.67
Societal Responsibility	0.45	0.60	0.53
Knowledge of Law/Government	0.37	0.37	0.37
Health	0.34	0.76	0.73
Creative Activities	0.95	0.96	0.94
Career Awareness	0.36	0.39	0.38
Appreciating Human Accomplishments	0.99	0.94	0.89
Knowledge of Human Accomplishments	0.34	0.35	0.33
Information Usage	0.36	0.39	0.38

Table 7

Median Values for Percentage of Omits-1978

Area	Grade 5 (N=457)	Grade 8 (N=151)	Grade 11 (N=128)
Self-Esteem	1.2%	0.4%	0.3%
Understanding Others	0.4	0.2	0.2
Reading	0.3	0.4	0.6
Writing Skills	0.3	0.3	0.3
Mathematics	1.2	0.9	1.0
Interest in School and Learning	0.3	0.4	0.3
Societal Responsibility	0.3	0.1	0.1
Knowledge of Law/Government	0.6	0.6	0.7
Health	0.2	0.3	0.1
Creative Activities	0.5	0.3	0.3
Career Awareness	0.3	0.4	0.2
Appreciating Human Accomplishments	0.4	0.3	0.1
Knowledge of Human Accomplishments	23.7[1]	21.4[1]	30.7[1]
Information Usage	0.2	0.3	0.3

[1]A large percentage of omits for Knowledge of Human Accomplishments derives from counting the "I don't know" response as a non-response.

VALIDITY

WHAT IS VALIDITY?

Validity is evidence that helps separate fact from fiction in test results. In the case of the Educational Quality Assessment program, information on the validity of each scale can help educators translate the *paper goals* and *number outcomes* into a better understanding of Pennsylvania school-age children. No single procedure or experimental design gives a complete picture of a test's validity. Instead, each study, in its own way, helps round out a picture of what each instrument measures and, therefore, helps interpret test outcomes.

DEVELOPMENT PROCEDURES

It should be pointed out that various procedures incorporated in instrument development are designed to enhance validity. Specifically, information derived from student interviews and from review committees are most helpful in designing content for both cognitive and noncognitive areas which are relevant to the curriculum and to the situations encountered by students at various grade levels.

VALIDITY STUDIES

Conducting validity studies requires a great deal of staff time: designing an appropriate study, making arrangements with schools willing to participate, preparing materials, going to schools to administer tests, preparing data, conducting statistical analyses, writing reports, and providing schools feedback. For these reasons it was not possible to conduct validity studies on each of the 42 instruments comprising the complete EQA package. The following summary provides a brief overview of the kind of validity evidence gathered for each area measured by the 1978 package:

Area	*Summary of Validity Studies Conducted on 1978 Assessment Package*
Self-Esteem	A study was conducted which related EQA scores to scores from the Self-Concept as a Learner scale (SCAL) and teacher ratings. In addition, the new EQA measure was correlated with the EQA instrument in use from 1974-77.
Understanding Others	A study was conducted which related EQA scores with criterion groups based on teacher ratings.
Reading	A study was conducted which correlated EQA reading scores with reading scores on the Stanford or Iowa achievement tests. In addition, a passage dependency study was conducted.

Writing Skills	A study was conducted which correlated EQA writing scores with language arts scores derived from the Stanford or Iowa achievement tests.
Mathematics	A study was conducted which correlated EQA mathematics scores with mathematics scores derived from the Stanford or Iowa achievement tests.
Interest in School and Learning	A study was conducted which related EQA scores to teacher ratings of five traits associated with interest in school and learning. EQA scores were also correlated with student self-assessment of behaviors relevant to interest in school.
Societal Responsibility	Since the measurement of this area did not undergo major change, no new validity study was conducted. It is presumed that the 1974 validity study which compared categories of institutionalized girls and noninstitutionalized girls is still appropriate.
Knowledge of Law/Government	No study was conducted.
Health	At grade 5, a study was conducted which correlated EQA scores with scores on the North Carolina State Assessment Test of health knowledge. At grades 8 and 11, a study was conducted which related EQA scores to health behavior criterion groups based on teacher observations.
Creative Activities	No study was conducted.
Career Awareness	No study was conducted.
Appreciating Human Accomplishments	No study was conducted.
Knowledge of Human Accomplishments	No study was conducted.
Information Usage	A study was conducted which correlated EQA scores with scores derived from the Cornell Critical Thinking Test and the Critical Thinking Test used in Montana's statewide assessment program. In addition, EQA's scores were related to criterion groups based on teacher ratings.

ABSTRACTS OF VALIDITY STUDIES CONDUCTED

In this section are brief abstracts of the validity studies conducted with the EQA instruments. Larger reports for most of the studies are available from the authors.

SELF-ESTEEM

Relationships Among Several Measures of Self-Esteem

Robert N. Reynolds
Division of Research

The validity of the new EQA measure of self-esteem was assessed by examining its relationship with several other measures of self-esteem.

More than 300 children in each of the three grades (5, 8 and 11) completed at one sitting an instrument package which included the new EQA measure, the former EQA self-esteem measure and the Self-Concept As a Learner scale . Independent teacher ratings of each child on two basic components of self-esteem, "feelings of acceptance" and "feelings of success," were gathered, according to a five-point rating scale. The resulting data were used in separate correlational analyses for each grade level.

The results of these analyses provide substantial evidence of the validity of the new EQA measure of self-esteem. At each of the three grade levels the relationship between total scores on this measure and the former EQA measure was quite strong (r's of .70, .80 and .69). A similar relationship was established with the Self-Concept As a Learner scale (r's of .76, .81 and .76). In addition, the analyses showed significant correlations between teacher ratings on the two components of self-esteem and the scores on the instrument.

UNDERSTANDING OTHERS

Validation of the Understanding Others Instruments

George Brehman
Division of Research

Evidence of the validity of the Understanding Others instrument was obtained by asking the teachers in the 5th, 8th and 11th grades of a volunteer school district to each list five of their students whom they considered to be high in their acceptance of those who are different from them, and five who they considered to be relatively low in this regard. The students so nominated were then placed in one of two criterion groups, i.e., a high acceptance or a low acceptance group.

At no time were the names of those nominated available to the experimenters. All students were assigned a number for the purpose of the study by the school and these numbers were used instead of names on both the answer sheets and for the nomination of high and low acceptance criterion group students.

The Understanding Others total scores and scores based on significant subsets of items reflecting the various categories of difference and types of situations in the instrument were then determined for students in each of the two criterion groups.

For grades 5 and 8 it was found that the high and low acceptance groups differed significantly from one another with regard to the total score on the test instrument. For grade 5 this difference, based upon item subset analysis, proved to be primarily due to significantly different scores on items dealing with religion and nationality (language) differences and superordination situations; superordination is a situation where a student works under the direction or authority of another student who is different from him or her.

For grade 8, the findings unequivocally supported the instrument's validity in that the 8th grade criterion groups differed significantly from one another with regard to the total score and those items dealing with socioeconomic, religious, racial, national (language) and appearance differences as well as items dealing with friendship, respect, social distance, and superordination situations.

No findings supporting the validity of the test instrument were found for grade 11. These results, however, were rendered uninterpretable due to several factors that may have reduced the likelihood of significant findings. It was found, in addition to the relative homogeneity of the district's student body, that the teachers were unreliable in their nominations of the senior high level. Several students were nominated for both the high and low criterion groups. It was also found that most of the 11th grade students in the "low" group gave across-the-board maximally favorable responses to the items suggesting a deliberate attempt to obtain a spuriously high score. In addition, there was considerable overcrowding the senior high testing situation.

Further efforts should be made to validate this instrument at the 11th grade, but it should be noted that reliability estimates and item characteristics (means, standard deviations, etc.) for all students were found to be essentially the same for all grade levels. This lends credence to the hypothesis that the problem lay with the validation process for grade 11 rather than with the instrument itself.

COMMUNICATION SKILLS AND MATHEMATICS

The Relationship of the EQA Reading, Writing and Mathematics Tests to Standardized Achievement Tests

Richard L. Kohr
Division of EQA

A frequently asked question is, "To what extent do the EQA basic skills tests relate to the commonly used standardized achievement tests?" The EQA reading, writing skills and mathematics tests were administered to students in various school districts that use the Iowa or Stanford and, in one case, the California Achievement Test. The available total and subscale scores on the commercial tests were correlated with scores on the appropriate EQA test. The results, presented in Table 8, show that most of the correlations are in the .70's and .80's.

In one district, the 30-item grade 5 math test used from 1970 through 1977 was correlated with the present 60-item test. The observed relationship of .75 indicates that students scoring high on the previous EQA math test also received high scores on the present test.

A passage dependency study was also conducted with the Reading test. This study is described in greater detail toward the end of this chapter.

INTEREST IN SCHOOL AND LEARNING

How Do Teacher Ratings of Student Behavior Relate to EQA Interest in School and Learning Scores?

James P. Dorwart
James R. Masters
Division of Research

The purpose of this study was to investigate the validity of the EQA Interest in School and Learning instruments. Fifth, 8th and 11th grade students from four school districts were given the EQA inventory and another questionnaire. The questionnaire asked students for further information about their behaviors relevant to interest in school and learning. These same students were rated by their teachers on five traits associated with interest in school and learning, i.e., desire to attend class, achievement relative to ability, relationship with teachers, class behavior, and interest in learning. The 8th and 11th grade students were rated by administrators or counselors with respect to discipline and truancy.

The results indicated support for the validity of the EQA instruments. The students rated high on interest-in-school and learning traits scored significantly higher on the EQA instruments than the students who were rated low on interest-in-school and learning traits. This significant difference was observed for 5th, 8th and 11th grade students. Each student's EQA instrument score was converted to a percentile indicating how he or she scored relative to all other students tested at that grade level. The average percentile ranking for the 5th grade students who were rated high on the interest-in-school and learning traits was 64.5 but only 39.3 for students who were rated low. The results for the grade 8 and 11 students were even more conclusive. The students rated high had average percentile rankings of 70.8 and 64.1 for grades 8 and 11 respectively, while the students rated low had average percentile rankings of 27.6 and 37.2 for grades 8 and 11 respectively. These results show that, on the average, those students who were rated high on the interest-in-school and learning traits by their teachers were the same students who scored high on interest-in-school and learning as measured by the EQA instruments.

Additional support for the validity of the instruments was found upon analysis of results of the student questionnaire. At grade 5 there was a positive, significant correlation between the EQA instrument scores and the number of subjects that the students felt were interesting. At grades 8 and 11 there were positive, significant correlations between EQA scores and 14 items on the student questionnaire. Among these were items which measured the students' assessment of how frequently they paid attention in class, studied outside of class, were willing to discuss problems with teachers and discussed school work with their parents.

SOCIETAL RESPONSIBILITY

Discrimination of Societal Responsibility Characteristics in an Institution for Neglected and Delinquent Girls

N. F. Russell
Division of EQA

This study was conducted in a home for wayward girls in western Pennsylvania. The institution has a policy of periodic appraisal (every three months) of each resident's social and emotional

adjustment. Evaluations are made by a three-member team consisting of a psychologist, a case worker and a cottage attendant. Characteristics evaluated include personal responsibility, honesty, fighting behaviors and ability to get along with others. Assignments to living quarters were made contingent upon the above evaluations. As a girl's adjustment improves, she is moved to another cottage with girls of similar adjustment and with more privileges. There are five cottage units in the institution; cottage 5 houses the most adjusted of the students.

The EQA societal responsibility scale (1974-77 version) was administered to the entire population of girls (N=46) in that institution. For purposes of analysis, the girls were separated into two groups. Those living in the first three cottages formed group one (most maladjusted, N=25). Group two (most adjusted, N=21) was composed of girls in cottages 4 and 5.

Analyses revealed that the scale could (1) discriminate between these two groups (the adjusted group scoring higher) and (2) discriminate between the total institutional group and a group of public school children having similar backgrounds (the public school group scoring higher).

Note: This study, conducted in 1974, is listed here because of its relevance to the present societal responsibility instruments used in grades 8 and 11.

HEALTH

Relationships Between the EQA Health Tests and Other Measures of Health Knowledge and Habits

Carol Bellew
Division of Research

Grade 5

The EQA grade five health knowledge test was validated by correlating EQA test scores with scores made by the same students on the health knowledge section of the North Carolina State Assessment Test. Since the aim of both tests was similar, a high degree of relationship would indicate support for the validity of the EQA test.

Both tests were administered to 286 students in the 5th grades of two school districts. The analyses of the data showed a high positive correlation between the total scores (r=.76). Significant correlations were also found for three-item groupings common to the two tests. These areas were identified as sociological health (r=.43), safety (r=.53) and personal health (r=.64).

Grades 8 and 11

For validation of the grade 8 and 11 health and safety practices test, teachers were asked to identify students with outstandingly high and outstandingly low health practices in four areas: personal health, safety, the use of drugs, alcohol and tobacco, and physical fitness. The ratings were then correlated with the students' scores on the EQA health and safety practices test.

Two hundred and seventy 8th grade students and 244 11th grade students of two school districts participated in the project.

The data analyses showed that students of both grades who were judged by teachers to have outstandingly good health habits scored significantly higher on the EQA health and safety practices instrument than students judged to have poor health habits. Significant differences between the two groups were also found for the drug, alcohol and tobacco, safety and physical fitness item groupings. For the personal health item grouping, no difference was found at grade 8. At grade 11, however, a significant difference was found when only students who received two or more ratings in this area were included in the analysis. Thus, in general, support was found for the validity of the health and safety practices instrument. Support was also found for item groupings measuring drug, alcohol and tobacco, safety and physical fitness practices.

INFORMATION USAGE

Relationships Between EQA Information Usage Scores and Measures of Critical Thinking

James R. Masters
James P. Dorwart
Division of Research

The EQA information usage instruments were intended to measure logical thinking of a type similar to that called for in a number of the nationally-known critical thinking tests. Thus, a major focus of this study was the examination of relationships between the EQA instruments and other accepted measures of critical thinking.

The study took place in three school districts at the grade 5 level and in two districts in both grades 8 and 11. A total of 182 grade 5 pupils, 197 grade 8 students and 208 grade 11 students participated.

Validity support was found for the instruments at all three grade levels. Strong relationships existed between scores obtained by the same students on the EQA instrument and on the Cornell Critical Thinking Test (r=.68 and .64 at grades 8 and 11, respectively). Correlations with the critical thinking test used in Montana's statewide assessment program were also high, .63, .62 and .73 at grades 5, 8 and 11 respectively.

As another assessment of validity, teachers were asked to rate the critical thinking abilities of the students tested by picking out those they believed to have a high level of this ability and those they believed to have a low level. At all three grade levels, students who were rated high in critical thinking ability scored significantly higher on the Information Usage tests than students rated low. Compared to all students tested at their grade levels, the average student chosen by teachers for the high group scored at the 79th percentile on the EQA test in grade 5, at the 68th percentile in grade 8 and at the 82nd percentile in grade 11. The average student chosen by teachers for the low group scored at the 25th percentile in grade 5, at the 30th percentile in grade 8 and at the 25th percentile in grade 11.

The study, supports the validities of the Information Usage instruments. The EQA instruments were found to correlate highly both with accepted critical thinking instruments and with teacher ratings of student critical thinking abilities.

READING PASSAGE DEPENDENCY STUDY

Some of the recent testing literature recommends that reading comprehension items be passage dependent, i.e., getting items correct depends on the context of the reading passage. For items which are passage dependent, a student must derive information or make a proper inference based on a careful reading of the passage and then correctly answer questions based on the material. If students can answer the questions correctly *without* reading the passage, the test hardly assesses the ability to comprehend written materials. In this case, a test could be called "passage independent" and most likely measures some characteristic such as knowledge already possessed. When students are unable to answer many questions correctly without reading the passage, the test would be considered passage dependent.

Some researchers such as Pyrczak (1972) defined passage dependency as the extent to which questions referring to a particular reading passage can be answered correctly in the absence of that passage. Pyrczak formulated an index for examining individual items which required only the administration of a reading test item with the passage *absent*. This approach is mainly concerned with comparing the number of items correct under the passage absent condition with the number of items expected to be correct by chance. Accordingly, the closer the mean test score is to the chance expectation level, the more passage dependent the test.

A different procedure was applied by Tuinman (1973), who administered test items with passages absent as well as with passages present. His approach was to compare the mean test score obtained under both conditions. Generally, the more the mean score under the passage present condition exceeds the passage absent score, the greater the passage dependency. The effort to gather passage dependency evidence for the EQA reading tests more closely follows the approach of Tuinman. Two special forms were prepared. In one, the items for *odd* numbered passages (1st, 3rd, 5th, etc.) were presented with the passages *present* and the items for *even* numbered passages (2nd, 4th, 6th, etc.) were presented with the passages *absent*. In the second, the items for *odd* numbered passages were presented with the passage *absent* and the item for *even* numbered passages were presented with the passages *present*.

All 48 of the EQA reading items for grades 8 and 11 were used in the passage dependency study. At grade 5, 47 of the 48 items were used. One item was eliminated because there was no discernable way of including it without cueing the correct answer to another item. Test booklets were prepared for administration by interspersing the forms so that when handed out, half the students would randomly receive each form. The test was administered by the teachers of the cooperating school districts. After the study was completed, it was found that the test distribution procedure was handled properly at grade 5, but a deviation took place at grades 8 and 11. At those grade levels, there was an apparent misunderstanding in which the two forms were separated and given to separate classes. This meant that the students receiving each form were not random groups.

In Table 9 are the results of the passage dependency study for each grade level. In the table "Form" refers to the specially constructed forms used in this study; "N" refers to the number of students tested; "No. Items" refers to the number of items presented with or without the passage; "Ave. Prop. Right" refers to the average proportion of items correct (obtained by dividing the number correct by the number of items); "KR-20" is the internal consistency estimate of reliability and "r" is the correlation between the scores on the set of items with the passage present and scores based on the set of items with the passage absent.

Support for passage dependency is found in that the average proportion correction for items with the passage present is greater than for items with the passage absent. This can be observed in two ways. First, a comparison may be made within the same form. Notice that for grade 5, Form 1, the value of .55 exceeds the value of .39. Although they represent different sets of items, the results are in the expected direction. Secondly, a comparison can be made across forms. To do this, compare the .55 on the 25 items (with passage present) on Form 1 with .32 observed on the same set of items (but with passage absent) on Form 2. Here again, there is a pattern of higher scores for those items with the passage. Similarly, one can compare results within Form 2 for grade 5. Here a value of .55 was also observed for the 22 items presented with a passage present which exceeds the value of .32 for the 25 items on the same form with passage absent. Also, one can see that on the same set of 22 items which appeared on Form 1, the average proportion correct was lower, only .39.

The trend at grade 5 can also be observed at grades 8 and 11, thereby supporting the passage dependency of the items. Table 1 shows that the KR-20 values are quite high for each of the passage present conditions. As one might expect, internal consistency is considerably lower when passages are absent and guessing increases.

The correlations between scores with and without passages are fairly substantial, despite the low reliability of passage absent scores. The only exception is at grade 11, Form 2, where the correlation was .12.

In summary, it could be said that the data in Table 9 supports the passage dependency of the items in the EQA reading test.

Table 8

Correlations Between EQA Basic Skills Instruments and Commonly Used Standardized Tests

EQA Area	Standardized Test	Grade 5		Grade 8	Grade 11
READING	Stanford				
	Vocabulary	.76	.61	.73	None
	Comprehension	.78	.84	.75	None
	Word Skills	.74	.58	None	None
	Total	.80	.80	.79	.80
	Iowa				
	Vocabulary		.75	.74	.62
	Reading		.84	.76	.58
	Total		None	None	.66
WRITING SKILLS	Stanford				
	Spelling	.71	.76	.60	None
	Language	.74	.82	.83	None
	Total English	None	None	None	.78
	Iowa				
	Spelling		.55	.65	.47
	Capitalization		.80	.73	None
	Punctuation		.65	.68	None
	Usage		.38	.60	.72
	Total Language		.73	.76	.71
	California				
	Language		None	None	.79
MATH	Stanford				
	Concepts	.71	.74	.80	None
	Computation	.84	.69	.75	None
	Application	.68	.73	.77	None
	Total	.83	.79	.82	.83
	Iowa				
	Concepts	.58	.81	.86	None
	Problems	.58	.74	.71	None
	Total	.62	.82	.83	.77 .78

NOTE: Results were obtained from more than one district for some standardized tests at grade 5 and for the Math-Iowa at grade 11.

Table 9

Summary of Passage Dependency Analyses for
EQA Reading Comprehension Tests

Grade Level	Form	N	Passage Present			Passage Absent			r
			No. Items	Ave. Prop. Right	KR-20	No. Items	Ave. Prop. Right	KR-20	
5	1	50	25	.55	.77	22	.39	.05	.52
	2	52	22	.55	.80	25	.32	.31	.31
8	1	107	23	.61	.84	25	.39	.44	.50
	2	102	25	.51	.82	23	.34	.60	.60
11	1	107	20	.66	.74	28	.43	.39	.57
	2	103	28	.53	.83	20	.32	.26	.12

REFERENCES

Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, New York: Academic Press, 1969.

Pyrczak, F. Objective evaluation of the quality of multiple-choice test items designed to measure comprehension of reading passages. *Reading Research Quarterly*, 1972, *8*, 62-71.

Shoemaker, D. M. *Principles and Procedures of Multiple Matrix Sampling*, Cambridge, MA: Ballinger, 1973.

Tuinman, J. J. Determine the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly*, 1973, *9*, 206-223.

APPENDIX

Individuals Participating on Test Review/Advisory Committees

The following individuals served on committees who advised or reviewed EQA instruments which became part of the present package. Their recommendations are deeply appreciated and were essential to the preparation of the present instrument package. However, this does not mean that they would agree completely with each item that appears in the instrument for which they provided input.

Self-Esteem
and
Interest in School and Learning

Patricia Carmo
Harrisburg School District

Elmetta Carter
Abington School District

Dean Chandler
Penn Hills School District

James Cordis
McKeesport School District

William Cornell
Pa. State Education Association

George Demshock
Pa. State Education Association

Margaret Furlong
Upper Dublin School District

Norman Furrer
Altoona School District

Diane Ginn
Chester Upland School District

Lenore Loerch
State College Area School District

Antoinette McGrath
Cannon-McMillan School District

Rex Miller
Dover Area School District

Doug Simpkins
Clearfield School District

Mary Lou Stefanko
North Hills School District

Understanding Others

A formal review committee was not established for this area. Instead, correspondence from concerned educators and reactions from PDE staff, including members of the Office of Equal Rights, were the basis for revising the Understanding Others instrument.

Mathematics

Genevieve Battisto
Pocono Mountain School District

Frederick H. Bell
Pittsburgh, Pa.

James J. Bonner
East Lansdowne, Pa.

Howard G. Bower
Pittsburgh Board of Public Ed.

Bruce C. Burt
West Chester Area S.D.

Richard Cox
University of Pittsburgh

Doris E. Creswell
PDE Math Advisor

Henry A. Diehl
Allentown School District

Mabel M. Elliott
West Chester State College

Ralph Heimer
Pennsylvania State University

Carl Heilman
PDE Math Adviser

Donald Hepp
Moon Area School District

Ralph Hunsberger
Central Bucks H.S. East

Andrew V. Kozak
Pennsylvania State University

Dorothea I. Peeler
Community College of Allegheny Co.

Mildred M. Reigh
Indiana University of Pa.

Jesse Rudnick
Temple University

Eleanora G. Sandblade
Warren School District

Barbara D. Smith
Peters Township School District

Fred R. Stewart
Neshaminy School District

June L. Trudnak
Bloomsburg State College

Cecil R. Trueblood
Pennsylvania State University

Mark Wiener
West Chester State College

Reading

Rita Bean
University of Pittsburgh

John A. Gates
Edinboro State College

Patricia Horton
Educational Improvement Center

Joyce Lee
Pennsylvania State University

Beatrice Levin
Philadelphia City School District

John Meehan
PDE

Greg Morris
Pittsburgh Board of Education

John E. Nagle
Capital Area IU #15

Bernice S. Nickols
Allegheny Intermediate Unit

Steve A. Pavlak
University of Pittsburgh

Jean M. Roach
Intermediate Unit #1

Jack Smith, Jr.
Hickory Township School District

Margaret M. Sponsellor
Bloomsburg State College

Peggy Stank
PDE

Wilhelmina Taylor
PDE

Writing Skills

Judith Bogert
Pennsylvania State University

Philip S. Boggio
Greater Latrobe School District

Melvin Haber
Pennsylvania State University

Frederick Harwood
Philadelphia School District

Eugene F. Iagnemma
Vandergrift, Pa.

Stephen Koziol
University of Pittsburgh

Marilyn Lye
Vandergrift, Pa.

Genevieve Peters
Harrisburg, Pa.

Raymond Stopper
Tredyffrin-Easttown School District

Audry Wiest
New Cumberland, Pa.

Ron Zeigler
Camp Hill, Pa.

Special Critique of Writing Skills Test

John Fremer
Educational Testing Service

Irene Reiter
Philadelphia School District

Societal Responsibility and Knowledge of Law/Government

Consultants

Cathie Cornbleth
University of Pittsburgh

Ronald Galbraith
Allegheny Intermediate Unit

Elizabeth Haller
PDE

Henry Johnson
Pennsylvania State University

Daniel Nichols
Allegheny Intermediate Unit

Robert Schell
PDE

Leo R. West
East Allegheny High School

Robert Wingert
PDE

Health

Grade 5

Content balance and test items reviewed by Douglas Boelhouwer, PDE.

Grades 8 and 11 Special Task Force

Douglas Boelhouwer
PDE

Joseph Carr
PDE

Chick Hungerford
PSAHPER

Barbara Lockhart
PSAHPER

Vernon Register
PDE

Carol Underwood
PSAHPER

Ed Wickland
PSAHPER

Ann Zerbe
PDE

PSAHPER refers to the Pennsylvania State Association of Health, Physical Education and Recreation.

Educators Evaluating Health and Safety Practices Items at Grade 8

Beverly J. Bartholomew
North Versailles, Pa.

Ritz Benner
Millerstown, Pa.

Kathy Biffel
Reading, Pa.

Alma W. Brethaver
Downingtown, Pa.

Daniel R. Curren, Jr.
Springfield, Pa.

Marlene Furnish
Tionesta, Pa.

Thomas E. Hall
Oreland, Pa.

Dale E. Hassinger
Greensburg, Pa.

Sallie Hook
Slippery Rock, Pa.

Minnie Kemp
Pottstown, Pa.

Marjorie J. LaBella
Duke Center, Pa.

Kathi Jo McCleary
Stewartstown, Pa.

William A. Nelson
Harrisburg Area Community College

Janice Ann Neumayer
Philadelphia, Pa.

Richard L. Pae
Clarion State College

Charlene Potochar
Hooversville, Pa.

Gene W. Plum
Fawn Grove, Pa.

Jo Ann Regimbal
Ambler, Pa.

Evelyn M. Robinson
Philadelphia, Pa.

Ruth Ann Tew
Levittown, Pa.

John Turner
Philadelphia, Pa.

Pamela Veltri
Pittsburgh, Pa.

Pamela L. Williams
Reading, Pa.

Davies U. Bahr
Boalsburg, Pa.

William F. Bearse
Bethlehem, Pa.

Joanne Becker
Philadelphia, Pa.

Thomas J. Bottorf
Monroeville, Pa.

Alfred L. Brodhag
Lansdale, Pa.

Olga P. Cawley
Pittsburgh, Pa.

Louis P. Costanzo
Dunmore, Pa.

Fred Dean
Monroeville, Pa.

Dennis Douds
East Stroudsburg State College

Eugene Fritz
Conestoga, Pa.

Rosalyn A. Greenlee
Coatesville, Pa.

Marianne L. Herrity
East Stroudsburg, Pa.

Marilyn R. Istvan
East Pittsburgh, Pa.

Jack Kist
Stroudsburg, Pa.

Nicholas Machupa
Slippery Rock, Pa.

David H. Neff
Millersville, Pa.

Irene J. Norton
Philadelphia, Pa.

Ms. Debbie Pajak
Aliquippa, Pa.

Kamara L. Perry
Philadelphia, Pa.

John Pfau
Mifflinburg, Pa.

Maria Rifenberg
Greenville, Pa.

Deborah Shaw
Mt. Pleasant, Pa.

Earl S. Shive
East Stroudsburg, Pa.

Samuel Virgilio
Radnor, Pa.

Kathleen Wilson
Sayre, Pa.

Career Awareness

Karen Cava
Allegheny Intermediate Unit

Sylvia Charp
Philadelphia School District

Michael Ciavarella
Shippensburg State College

Cathy Lee Dzubak
Hopewell High School

Edwin L. Herr
Pennsylvania State University

Theodore Hoffman
PDE

Jerome T. Kapes
Lehigh University

Thomas Long
Altoona, Pa.

Isaiah McCloskey
Columbia Montour AVTS

Bruce McGlothlin
Allegheny Intermediate Unit

John Meerbach
PDE

David Minnis
Northwest Tri-County IU

Carl Pepperman
Central Susquehanna Valley IU

Jean Purvis
State Advisory Council for
Vocational Education

Arthur Schwartz
Department of Labor and Industry

Richard Stabler
Central Columbia High School

Creative Activities, Appreciating Human Accomplishments and Knowledge of Human Accomplishments

The following individuals from the Bureau of Curriculum Services, PDE, offered recommendations for these three areas.

William Bolles
Douglas Boelhouwer
Joseph Carr
Elizabeth Haller
Joseph McCarthy
John McDermott
Clyde McGeary
John Meehan
Vernon Register
Robert Revicki
Bernice Ruehner
Robert Schell
Robert Wingert

Information Usage

The complete development of this instrument was carried out principally by Division of Research staff with assistance by EQA staff. During development, helpful suggestions were made by several authorities, but at no time was a formal review committee convened. John Fremer of Educational Testing Service provided a critique of the instrument.